INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



CALIDOSCOPIOS

Febrero 1986

450 Ptas.
(IVA incluido)

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 113
Febrero 1986



EDITA Prensa Científica, S.A.
Calabria, 235-239
08029 Barcelona (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

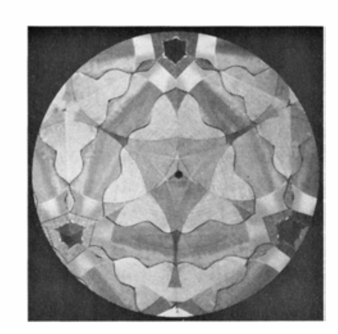
Josep-Enric Llebot: *Soporte lógico para la defensa antimisil*; Amando García Rodríguez: *Conjugación de fase óptica*; Manuel Crespo: *Comunicación por vía acústica en el grillo*; Isabel García Acha y Julio Rodríguez Villanueva: *SIDA y sistema inmunitario*; Luis Bou: *El teorema enorme* y *Juegos de Ordenador*; Mireia Bofill: *La alimentación en China*; Antonio Blanco Freijeiro: *Planos del templo de Apolo en Dídyma*; J. Vilardell: *Taller y laboratorio.*

Ciencia y sociedad:

Ricardo Génova y Juan Rius-Camps

Libros:

Carlos Pérez, Andrés de Haro, Antonio Lafuente y M. San Miguel



PORTADA

La fotografía de la portada muestra una imagen generada en un calidoscopio de polarización. Los trocitos de plástico transparente diferentemente tensados que hay en el extremo del tubo del calidoscopio opuesto al ocular cambian la polarización de la luz que penetra en el instrumento atravesando un filtro polarizador. En el extremo correspondiente al ocular hay un segundo filtro polarizador. Cuando se gira el tubo, se presentan a la vista diferentes trozos de plástico tensado. Como cada trozo cambia la polarización de la luz de un modo característico, parte de los colores del espectro luminoso atraviesan el segundo filtro polarizador y parte no. El observador contempla así una cambiante figura coloreada. En otros calidoscopios modernos se combina un diseño imaginativo con la técnica electrónica. En un modelo, los impulsos sonoros gobiernan un conjunto de diodos emisores de luz, de tal modo que la figura reluce y varía reaccionando a los sonidos (véase "Taller y laboratorio" en este mismo número).

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
08029 Barcelona (España)
Teléfono 322 05 51 ext. 33-37

Condiciones de suscripción:
España:
Un año (12 números):
4400 pesetas (IVA incluido)
Extranjero:
Un año (12 números): 33 U.S. $
Ejemplar atrasado ordinario:
450 pesetas (IVA incluido)
Ejemplar atrasado extraordinario:
575 pesetas (IVA incluido)

Distribución para España:
Distribuciones de Enlace, S. A.
Bruch, 49 - 08009 Barcelona

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - 08029 Barcelona

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137 - 28030 Madrid
Teléfonos 430 84 81 - 437 12 91
Cataluña:
Paulino Muñoz Victoria
Comunicación Diaria, S.A.
Aribau, 195, 4.° E - 08021 Barcelona
Teléfono 200 17 58

Controlado por O.J.D.

PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la portada de Quesada/Burke

Página	Fuente	Página	Fuente
11-16	Hank Iken, Walken Graphics	53	Luc Montagnier, Instituto Pasteur, París
17	U.S. Navy	54	Ilil Arbel
21	A. V. Mamaev y N. A. Melnikov	57-58	Roser Gonzàlez-Duarte, Elvira Juan, Lluïsa Vilageliu, Sílvia Atrian y Ricardo Génova
22	Ian Worpole	71	Gabor Kiss
23	A. V. Mamaev (*arriba izquierda*), Ian Worpole (*arriba derecha y abajo*)	72-73	Edward Bell
24-26	Ian Worpole	74	Gabor Kiss
28-29	Theo Weber, Instituto Max Planck de Fisiología de la Conducta, Seewiesen, República Federal de Alemania	75	Edward Bell
		76-78	Gabor Kiss
		80-81	Edward Bell
		85-86	Vaclav Smil, Manitoba
		88-92	Joan Starwood
30-36	Patricia J. Wynne	95	Lothar Haselberger (*arriba*), Raymond V. Schoder, S. J. (*abajo*)
38-39	Ricardo Génova		
41	Juan Rius-Camps		
44	Ilil Arbel	96	Alan D. Iselin
45	George Janossy, Royal Free Hospital, Hampstead, Inglaterra	97	George Niemann (*arriba*), Alan D. Iselin (*abajo*)
		98	Alan D. Iselin (*izquierda*), Hans Georg Bankel (*derecha*)
46	Jeffrey Laurence y Powers Peterson, Facultad de Medicina de Cornell (*arriba*); Ira Schierem, Facultad de Medicina de Cornell (*abajo*)	99	Alan D. Iselin
		100	Lothar Haselberger
		101-102	Alan D. Iselin
		105-107	Andrew Christie
		110	Quesada/Burke
47-52	Ilil Arbel	111-115	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa, S.A.
Pedro IV, 160 - 08005 Barcelona
Fotocromos reproducidos por GINSA, S.A.
Gráfica Elzeviriana, S.A.
Nápoles, 249 - Tel. 207 40 11
08013 Barcelona
Printed in Spain - Impreso en España

Los autores

HERBERT LIN ("Soporte lógico para la defensa antimisil") desarrolla estudios de ampliación postdoctoral en el Centro de Estudios Internacionales del Instituto de Tecnología de Massachusetts (MIT). Se licenció y se doctoró en física por el MIT en 1973 y 1979, respectivamente. Aquí enseñó antes de pasar a la Universidad de Washington y, luego, a la de Cornell. En 1984 regresó al MIT.

VLADIMIR V. SHKUNOV y BORIS YA. ZEL'DOVICH ("Conjugación de fase óptica") trabajan en el Instituto de Problemas de Mecánica de la Academia de Ciencias de la Unión Soviética. Shkunov se educó en el Instituto Físico-Técnico de Moscú. Zel'dovich, que se licenció por la Universidad estatal de Moscú en 1966, prosiguió su formación académica en el Instituto de Física Teórica y Experimental de Moscú. Pasó en el Instituto P. N. Levedev de Física de Moscú, donde se doctoró en ciencias en 1980. Zel'dovich y un grupo de colegas fueron distinguidos con el Premio Estatal de la Unión Soviética por su investigación sobre la conjugación de fase óptica.

FRANZ HUBER y JOHN THORSON ("Comunicación por vía acústica en el grillo") son, respectivamente, director de departamento y asesor del Instituto Max Planck de Fisiología del Comportamiento de Seewiesen, Alemania Occidental. Huber se licenció por la Universidad de Munich, donde también se doctoró, en 1953, con una tesis sobre el sistema nervioso del grillo. De 1954 a 1960 fue profesor auxiliar de fisiología animal de la Universidad de Tübingen. La de Colonia le nombró en 1963 profesor de zoología y fisiología animal. Thorson se licenció en física (1958) por el Instituto Politécnico Rensselaer. Contratado por la General Electric Company, trabajó en el dominio de la biofísica. Su experiencia en ese centro le animó a ampliar estudios de biología en Los Angeles, doctorándse en zoología en 1965. De 1967 a 1969, ejerció de profesor auxiliar e investigador en la Universidad de California en San Diego. Ha colaborado en investigaciones sobre visión humana y de insectos, contracción muscular, transformación sensorial y comportamiento de insectos.

JEFFREY LAURENCE ("SIDA y sistema inmunitario") enseña medicina en el Hospital Clínico de la Universidad de Cornell y trabaja en el Hospital de Nueva York. Tras cursar el primer ciclo de carrera en la Universidad de Columbia se licenció en medicina por la Facultad Pritzker de la Universidad de Chicago. Siendo estudiante disfrutó de una beca Rhodes y otra de la Fundación Henry Luce. Dedicó un año entero a la investigación sobre las defensas inmunitarias contra el crecimiento tumoral en el Instituto Oncológico de la Universidad de Osaka. De regreso a los Estados Unidos completó su formación médica en el Hospital de Nueva York y se entregó a la investigación en la Universidad Rockefeller. Se incorporó al claustro docente de Cornell en 1982.

R. GONZALEZ-DUARTE, E. JUAN, LL. VILAGELIU y S. ATRIAN ("El sistema alcoholdeshidrogenasa"), profesoras del departamento de genética de la Facultad de Biología de la Universidad de Barcelona, han trabajado conjuntamente durante varios años en diversos aspectos relacionados con la alcoholdeshidrogenasa (ADH) de *Drosophila*. Gonzàlez-Duarte se doctoró en ciencias biológicas por la Universidad de Barcelona en 1972. Disfrutó de una beca del Consejo Británico para realizar una estancia de dos años en el departamento de biología molecular de la Universidad de Edimburgo, donde abordó algunos aspectos de la ADH en *D. melanogaster*. En el año 1974 inició en el departamento de genética de Barcelona una línea de investigación sobre esta enzima en diversas especies del género. Juan, doctorada en ciencias biológicas por la Universidad de Barcelona en el año 1980, purificó y caracterizó la enzima ADH en las especies *D. simulans, D. virilis* y *D. lebanonensis*. Con una beca estuvo en la Universidad de Edimburgo realizando estudios estructurales de la citada enzima. Vilageliu se doctoró en ciencias biológicas en 1981 por la Universidad de Barcelona. Sus trabajos sirvieron para profundizar en el estudio bioquímico de la ADH de *D. funebris*y *D. immigrans* y desentrañar su función en el metabolismo de los alcoholes. Atrian obtuvo una beca del Ministerio español de Educación y Ciencia para investigar la estructura de la ADH en *D. hydei*. Se doctoró en ciencias biológicas por la Universidad de Barcelona hace un par de años 1984.

DANIEL GORENSTEIN ("El teorema enorme") ocupa la cátedra Jacqueline B. Lewis de Matemáticas de la Universidad de Rutgers. Se educó en la Universidad de Harvard, por la que se licenció en 1943 y se doctoró en 1950. Comenzó su carrera como profesor auxiliar en la Universidad Clark, en la que llegó a profesor titular en 1959. En 1964 se trasladó a la Universidad del Nordeste. Tras pasar un año en el Instituto de Estudios Avanzados de Princeton, ingresó en Rutgers en 1969. En 1972 fue nombrado miembro de número de la Fundación Guggenheim y académico Fulbright; en 1978 prestó servicios como Profesor Distinguido en la Cátedra Sherman Fairchild del Instituto de Tecnología de California. Dedicado al estudio de los grupos finitos desde 1960, ha desempeñado un papel clave en el desarrollo del teorema de clasificación de todos los grupos finitos simples. En la actualidad, Gorenstein elabora una demostración más concisa de "segunda generación".

VACLAV SMIL ("La alimentación en China") enseña geografía en la Universidad de Manitoba en Winnipeg. Tras su licenciatura por la Universidad Carolina de Praga, trabajó en una oficina de planificación regional, como asesor en temas energéticos y del medio ambiente. A raíz de la invasión soviética de Checoslovaquia, en 1968, se trasladó a los Estados Unidos. En 1972 se recibió de doctor por la Universidad estatal de Pennsylvania y, posteriormente, se incorporó al claustro de profesores de la Universidad de Manitoba. Desde sus tiempos de estudiante en Praga, Smil ha centrado su interés en las interrelaciones entre producción alimentaria, energía y transformación del medio ambiente.

LOTHAR HASELBERGER ("Planos para la construcción del templo de Apolo en Dídyma") es un arqueólogo especializado en arquitectura helenística. Estudió arquitectura, urbanismo e historia de la arquitectura en la Universidad Técnica de Munich y en la de Harvard. Bajo el patrocinio de la Universidad Técnica de Munich pasó dos años investigando una clase poco usual de torres antiguas que se encuentran en las islas de Grecia; recientemente ha obtenido el grado de doctor por aquel trabajo. Desde 1980, con la ayuda del Instituto Arqueológico Germánico y Consejo de Investigaciones Científicas alemán, Haselberger ha estado transcribiendo y catalogando los planos de construcción descubiertos por él en Dídyma.

Hace…

José María López Piñero

… doscientos años

Francesc Salvá i Campillo ingresó en la Real Academia de Ciencias Naturales y Artes de Barcelona.

Había nacido en 1751 en la misma Barcelona y era hijo de Jeroni Salvá i Pontich, médico del Hospital de la Santa Creu. Recibió su formación primaria y secundaria, desde 1757 a 1766, en el Colegio Episcopal de la capital catalana. Estudió a continuación tres años de medicina en la Universidad de Valencia, presentándose después a las pruebas de grado en la de Huesca, donde obtuvo en 1771 el título de bachiller en medicina. Consiguió luego el de doctor en la Universidad de Toulouse y lo revalidó también en Huesca. De regreso a Barcelona, perfeccionó su preparación clínica trabajando junto a su padre en el Hospital de la Santa Creu y se convirtió muy pronto en uno de los médicos más prestigiosos de la ciudad. En 1773 ingresó en la Real Academia Médico-Práctica y en 1786, como hemos adelantado, en la de Ciencias Naturales, instituciones que sirvieron de marco a su actividad científica hasta su muerte en 1828.

Cuando estudió en la Universidad de Valencia, Salvá utilizó como textos los libros de Andrés Piquer que, junto a los *Commentaria* de Gerard van Swieten a los aforismos de Hermann Boerhaave, constituyeron la base inicial de su formación médica. No resulta extraño que su obra madura fuera después uno de los principales exponentes de la mentalidad antisistemática en la etapa final de la medicina española de la Ilustración. Sus diferencias respecto a Piquer fueron muy importantes, porque Salvá fue un médico rigurosamente al día, que dominó los saberes y las técnicas propias de los años de tránsito del siglo XVIII al XIX. Por ello, en su obra, desapareció el peso decisivo que la medicina tradicional había tenido en la labor de Piquer, hasta el punto de que, como vamos a ver, rechazó explícitamente el modelo hipocrático.

La influencia directa de la “Alte Wiener Schule” en la obra juvenil de Salvá le condujo, en primer término, a redactar un compendio en dos volúmenes de los *Commentaria* de Van Swieten, que no llegó a imprimirse. Sus primeras publicaciones fueron dos libros en defensa de la inoculación antivariólica (1777), destinados a contradecir los ataques que contra esta técnica preventiva había dirigido Anton de Haen, otro destacado miembro de la escuela de Viena por el que Salvá no ocultó, por otra parte, su admiración.

La viruela fue un tema que continuó ocupando a Salvá durante las siguientes décadas. Le dedicó varios trabajos, entre los que destaca una memoria que premió en 1790 la Société Royale de Médecine de Paris, y que su autor publicó, ocho años después, en las *Memorias* de la Real Academia barcelonesa, con el título de *Sobre las utilidades y daños de los purgantes y de la ventilación de las viruelas*. Tras el descubrimiento de la vacunación antivariólica por Edward Jenner, Salvá figuró entre los primeros médicos españoles que aceptaron el nuevo procedimiento. Contribuyó decisivamente a su difusión en Cataluña y fue también uno de los corresponsales que en torno al tema tuvo Ignacio María Ruiz de Luzuriaga, al que remitió “pus vacuno”.

Por el contrario, Salvá mantuvo una actitud escéptica frente al preparado antifebril de José Masdevall, que había conseguido gran celebridad al ser aplicado en diferentes epidemias. El volumen en el que analizó sus ventajas y limitaciones (1790) figura entre sus mejores trabajos médicos publicados durante los años finales de siglo. Junto a él sobresalen una memoria sobre el escorbuto y la fiebre tifoidea (1794), que redactó con motivo de uno de los concursos convocados por la Société Royale de Médecine, y un estudio acerca de las epidemias que había observado en el departamento de mujeres del Real Hospicio de Barcelona (1798).

A partir de 1765, la Academia Médico-Práctica había solicitado en varias ocasiones la creación en Barcelona de una cátedra de clínica. Gracias a la gestión de Salvá, se consiguió por fin, en 1797, la fundación de un Real Estudio de Medicina Práctica bajo la dirección de la Academia. Fueron nombrados catedráticos el propio Salvá y Vicent Mitjavila, pronunciando el primero la lección inaugural en 1801. Su obra médica madura la recogió fundamentalmente en dos volúmenes (1802-1818) que consagró a exponer su labor en dicho Real Estudio.

En los citados volúmenes defendió la enseñanza clínica desde una perspectiva típica de la mentalidad antisistemática. Van Swieten y, todavía más, Boerhaave, eran ya, para el Salvá de estos años, autores “tradicionales”. Le interesaban de modo más inmediato otras figuras, entre ellas, los miembros más recientes de la “Alte Wiener Schule”. Dedicó una amplia exposición a la semiología del pulso, basada en su recuento con un “relox de segundos”, en la que citó repetidas veces a Anton de Haen, cuyo interés por medir la realidad clínica compartía plenamente. Aceptó también la crítica que De Haen había hecho de la doctrina esfigmológica de Francisco Solano de Luque, porque juzgaba que el vienés estaba “más apoyado en la razón y la experiencia”. Le interesaron asimismo los criterios terapéuticos de Maximilian Stoll y, sobre todo, la obra patológica de Johann Peter Frank, cuya orientación general suscribía. Adoptó una actitud similar a la de este último ante la nosotaxia “more botanico”, considerando con escepticismo todas las clasificaciones, desde la defendida por François Boissier de Sauvages hasta la propugnada por Philippe Pinel. En este contexto propuso una reforma radical de la terminología médica (1807), limitada a la combinación de radicales griegos, que no tuvo repercusión alguna, lo mismo que otras iniciativas parecidas que en esta época se tomaron en diversos países europeos.

La obra y el pensamiento de Giovanni Batista Morgagni influyeron también notablemente en la producción médica madura de Salvá, cuyo fundamento más sólido fueron historias clínicas con cuidadosos protocolos de autopsia. Sin embargo, no llegó a incorporar plenamente los nuevos supuestos teóricos de la escuela anatomoclínica de París, tal como fueron formulados por Marie François Xavier Bichat, Jean Nicolas Corvisart y René Théophile Hyacinthe Laennec. En cambio, rompió abiertamente con el último reducto de la autoridad tradicional de los clá-

sicos, no admitiendo ya que las historias clínicas hipocráticas fueran un modelo adecuado para la práctica médica.

Dejando aparte otros aspectos de la obra de Salvá como médico, recordaremos su oposición a la unificación profesional de los médicos de formación universitaria y los cirujanos procedentes de los nuevos colegios creados por la reforma de la enseñanza quirúrgica. En un libro sobre el tema (1812), atacó ferozmente a las cabezas de dicha reforma: "El Colegio de Cirugía de Cádiz fue un patrimonio para los parientes de su principal fundador Perchet. El de Barcelona, para los de Virgili. El de Madrid, para los ahijados de Gimbernat y, mientras éste le gobernó solo, costó al erario más de un millón de reales sin haber habilitado un cirujano para la nación".

Salvá cultivó otras áreas científicas y técnicas. En colaboración con Francesc Sanponts i Roca, médico amigo suyo y autor de notables aportaciones a la ingeniería, ideó una agramadora cuya descripción fue publicada en una *Disertación sobre la explicación y uso de una nueva máquina para agramar cáñamos y linos* (1784). Se interesó por la navegación submarina, realizando algunas experiencias para solucionar el problema del aire respirable en el interior de un "barco-pez", y también por la aerostática. Su orientación ambientalista le condujo a realizar observaciones meteorológicas de modo sistemático a lo largo de casi medio siglo y a ocuparse de la construcción y perfeccionamiento de barómetros, higrómetros y otros aparatos. No obstante, su principal contribución extramédica corresponde a la electricidad y, en concreto, a la telegrafía eléctrica, campo del que es universalmente reconocido como un importante adelantado. Entre 1795 y 1804 leyó en la Academia de Ciencias Naturales y Artes tres memorias sobre el tema, que no fueron impresas entonces, sino publicadas por dicha institución en 1876 junto a unas "Noticias... sobre la invención de los telégrafos eléctricos debida al Dr. D. Francisco Salvá" redactadas por su discípulo Félix Janer.

La primera de dichas memorias se titula *Sobre la electricidad aplicada a la telegrafía* y fue presentada por Salvá a la Academia el 16 de diciembre de 1795. En ella propuso un telégrafo basado en descargas de botellas de Leyden, utilizando un alambre conductor para cada letra del alfabeto. Describió sus experiencias con un prototipo que había construido y se planteó los problemas prácticos que implicaba el tendido de las líneas: "Parecerá poco menos que imposible hacer correr desde lugares puestos a tres o cuatro leguas de distancia tantos alambres como se necesitan para hacer hablar la electricidad; pues aun cuando se hiciesen sostener por árboles muy altos, como entenas, que tuviesen un travesaño, en el que colocarlos separadamente o aislarlos, quedaban muy expuestos a que los muchachos los desbaratasen. Pero no es necesario que los alambres corran separados; pueden ir juntos a modo de una cuerda, sin que por esto vaya la electricidad por otros que por los que se requieren para el intento; y está claro que una cuerda de 44 alambres no está muy expuesta a que la rompan o

PROCESO
DE LA INOCULACION,
presentado
Al Tribunal de los Sabios,
para que la juzguen.
RESUMEN
DEL DICHO PROCESO
En un razonamiento dirigido á un Padre, que duda, si inoculará à su unico hijo.
POR EL DOCTOR EN MEDICINA
FRANCISCO SALVÁ,
I CAMPILLO,
Socio de la Academia Medico-Practica, establecida en Barcelona.

Con licencia. En Barc. Por Francisco Genèras, Bajada de la Carcel.
Vendese en Casa Francisco Ribas, Plaza de S. Jaime.

1. Portada de uno de los libros de Salvá en defensa de la inoculación antivariólica (1777).

(45)
ADVERTENCIA.
Las dos observaciones que siguen, pueden hacer ver, lo que fué la fiebre del puerto de Barcelona en 1803; y sirvieron para las lecciones sobre la fiebre llamada amarilla, que se dieron en el curso de medicina Clinica de 1804 á 1805, y por esto se insertan aqui.

OBSERVACION DE LA ENFERMEDAD
DEL CAPITAN OLANDÉS PLOUEN.

Circunstancias.

Plouen capitan de un barco Olandés, sugeto mayor de 60 años, afligido de profunda melancolia, con motivo de estarse comiendo en este puerto sus pocos caudales, por causa de la actual guerra con los Ingleses, habia venido á dicho puerto nueve meses atras en derechura desde Amsterdam, sin haber hecho otro viage que desde aqui á Nápoles, de cuyo reyno regresó, hace quatro meses, cargado de madera. Por otra parte, sobresaltado de la pronta y violenta muerte de algunos compañeros suyos, que parece murieron envenenados del cardenillo, el miércoles dia 19 de los corrientes mes y año se atracó de bacalao y judías, la misma tarde tubo calofrios, le entró calentura con muchas ganas de provocar, dieronle un vomitivo de quatro granos de tartrito estiviado de potaza, que le hizo vomitar excesivamente, y es lo que pudo averiguar el Doctor Don Buenaventura Vila Medico de cabezera, por medio del Interprete,

2. Página inicial de una de las "observaciones" utilizadas por Salvá para la enseñanza clínica en el Real Estudio.

MEMORIA

SOBRE

LA ELECTRICIDAD APLICADA Á LA TELEGRAFÍA.

EXCMO. SR.

La electricidad es de las ciencias con que puede demostrarse mejor, que los experimentos, que parecen más despreciables, han sido la base de los descubrimientos más importantes, y por consiguiente, puede animar más á los físicos á multiplicarlos. Con efecto ¿qué cosa tan indiferente á los ojos de muchos es, el que una punta de metal, puesta á cierta distancia de un conductor electrizado, se vea vestida de luz ó reluciente, y que este punto luminoso deje de encontrarse en una bola de hierro colocada al lado de dicha punta? Parece más propio para entretener á un niño, que para ocupar á un sábio, el que un cuerpo metálico puntiagudo chupe la electricidad á un conductor quietamente ó sin estrépito, y que se observe estallido cuando el flúido eléctrico pasa del conductor al globo de metal. Sin embargo, de los experimentos mencionados, el famoso médico Franklin sacó la atrevida idea de robar á las nubes impunemente la materia de los rayos y conducirlos á su albedrío á parajes en que no puedan causar daño. El golpe eléctrico de la portentosa botella de Leyden parece más propio para sorprender y encantar las

2

3. Página inicial de la primera de las memorias de Salvá sobre la electricidad aplicada a la telegrafía (1795), en la edición que publicó en 1876 la Real Academia de Ciencias Naturales y Artes, de Barcelona.

desbaraten los chiquillos, especialmente haciéndola sostener por mástiles bien altos. En las primeras pruebas que hice con un pequeño telégrafo medio armado, vestí los alambres de papel, después los rollé y siempre dirigí la electridad por los que quise. El papel barnizado con pez o con otra materia ideoeléctrica sería más propósito aún. Fuera de que, saliendo bien del modo insinuado, la cuerda de alambres podría correr por caños subterráneos, dividiéndolos para mayor precaución con una o dos líneas de alguna resina propia para aislar y así quedaba a descubierto de que las gentes la echasen a perder".

La segunda memoria, *Disertación sobre el galvanismo* (febrero de 1800), la dedicó a exponer una serie de experimentos en torno a la naturaleza del mismo que le llevaron a disentir de las teorías de Galvani y de Volta y a defender la propugnada por Alexander von Humboldt; incluía, además, una "Adición sobre la aplicación del galvanismo a la telegrafía", en la que describió otro telégrafo de su invención que utilizaba como señales las contracciones de músculos de ranas. Por último, en su *Memoria segunda sobre el galvanismo aplicado a la telegrafía* (febrero de 1804) estudió la utilización de la pila de Volta como generador de energía eléctrica para la transmisión de señales, siendo el primer científico que planteó la aplicación de este revolucionario aparato a la telegrafía.

Soporte lógico para la defensa antimisil

El proyecto de defensa denominado "guerra de las galaxias" se apoyaría en el control que ejercieran los ordenadores sobre una compleja red de armamentos. Quizá resulte imposible el desarrollo de programas fiables para tal sistema

Herbert Lin

"El objetivo final de la Iniciativa de Defensa Estratégica (IDE) es eliminar la amenaza planteada por los misiles balísticos nucleares" se afirma en la carta provisional del Organismo para la Iniciativa de Defensa Estratégica. Para alcanzar ese objetivo de defensa total, el Departamento de Defensa de los Estados Unidos ha autorizado a esa instancia la gestión de programas de investigación que examinen la viabilidad del desarrollo tecnológico necesario para una defensa contra misiles balísticos (BMD) destinada a proteger ciudades y asentamientos militares. Tal sistema defensivo destruiría o incapacitaría los vehículos de transporte de las ojivas nucleares, en el camino hacia sus objetivos, valiéndose de una amplia gama de armas defensivas: láseres, haces de partículas, guías electromagnéticas o vehículos no explosivos de impacto directo.

El Organismo para la Iniciativa de Defensa Estratégica reconoce que la importancia de la tecnología informática que controlaría cada una de las armas, y que coordinaría su funcionamiento, es comparable con la del desarrollo de las nuevas tecnologías de interceptación. El sistema de mando y control de la defensa total contra misiles balísticos debe ser capaz de recibir y actuar sin problemas con la información procedente de miles de rampas de lanzamiento, decenas de miles de cabezas nucleares y cientos de miles de señuelos. Debe hacerlo en la media hora que emplearía un misil balístico intercontinental (ICBM) para viajar desde su rampa de lanzamiento, en la Unión Soviética, hasta su objetivo en los Estados Unidos. Dado que en su mayor parte el sistema ha de estar automatizado, no habría tiempo para que la intervención humana corrigiera fallos inesperados. Una vez puesto en servicio un sistema BMD, las decisiones humanas se substituirían por la ejecución de un programa de ordenador.

Antes de que los Estados Unidos emprendan un serio esfuerzo en el desarrollo del soporte lógico de un sistema de tales características convendría plantearse tres cuestiones: ¿Cuál es la naturaleza de un sistema BMD? ¿Cuáles son los obstáculos que se opondrían al desarrollo de la programación del BMD? ¿Pueden salvarse esos obstáculos?

El sistema de defensa total que los funcionarios del Organismo para la Iniciativa de Defensa Estratégica (y sus críticos) han examinado más asiduamente consta de cuatro escalones, es decir, la defensa contra los misiles balísticos atacaría a los misiles hostiles en cada una de las cuatro fases de su vuelo, que son: la de lanzamiento, en la que un vehículo de varias fases sitúa la carga por encima de la atmósfera; la de postlanzamiento, durante la cual un vehículo maniobrable, el "autobús", libera secuencialmente por encima de la atmósfera terrestre vehículos de reentrada con ojivas nucleares y "ayudas a la penetración", como señuelos o trampas; la etapa intermedia, en la que los vehículos de reentrada y los señuelos recorren la mayor parte de su trayectoria, y la etapa terminal, en la que las cabezas nucleares a bordo de sus vehículos de transporte penetran en la atmósfera y detonan sobre sus objetivos.

En un sistema defensivo contra misiles balísticos desplegado en cuatro escalones, los interceptores disfrutan de varias oportunidades para destruir las armas ofensivas. En cada escalón, el sistema debe detectar y rastrear con éxito sus objetivos para llegar a destruirlos [véase "Defensa espacial y misiles balísticos", por Hans A. Bethe, Richard L. Garwin, Kurt Gottfried y Henry W. Kendall; INVESTIGACIÓN Y CIENCIA, diciembre de 1984]. Para coordinar la defensa y evaluar su eficacia se necesitan ordenadores y programas apropiados. Este proceso de coordinación se denomina dirección del combate. Aunque los expertos no se han decidido aún sobre cómo debería organizarse, una hipotética estructura comprendería ordenadores "locales" y programas responsables de la dirección del combate en cada escalón defensivo. Los sistemas de cada escalón estarían interconectados entre sí mediante un sistema global de dirección del combate.

Los programas que cumplieran la función de dirección dentro de cada escalón defensivo controlarían los sensores y las armas asignados al mismo. Tales sensores localizarían y rastrearían los objetivos potenciales y separarían los objetivos reales de los señuelos. Los programas crearían un "fichero de seguimiento" que contendría toda la información conocida acerca de cada objetivo. Asimismo, asignarían recursos defensivos en cada escalón específico, coordinando la información del fichero de seguimiento con el armamento disponible y con las reglas de acción, previamente programadas, que determinarían bajo qué circunstancias debería atacarse cada uno de los objetivos. El sistema global director del combate tendría en cuenta la envergadura y la naturaleza del ataque en curso y asignaría reglas de acción a cada escalón. Para preparar un sistema local de dirección del combate que se enfrentara a las cabezas nucleares que hubieran escapado de escalones precedentes, el sistema global debería trans-

mitir la información del fichero de seguimiento y de los sensores, obtenida en la fase de seguimiento, al escalón defensivo subsiguiente.

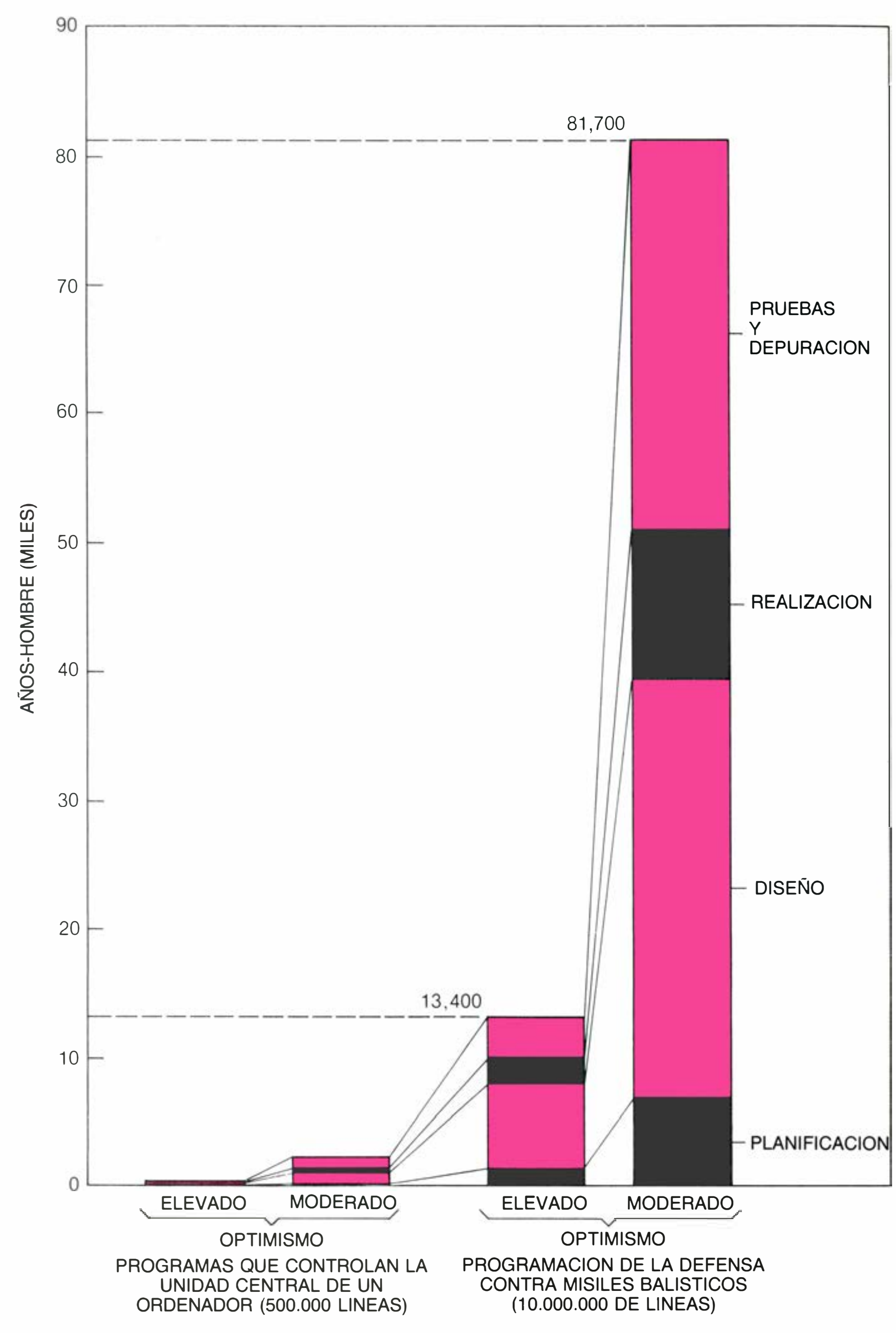
La defensa contra misiles balísticos dependería estrechamente de los programas que la controlan; cualquier programación defectuosa se traduciría en fracaso. Por tanto, el desarrollo del sistema de programación resulta un factor decisivo en el logro de los objetivos para los cuales se ha diseñado el sistema.

El desarrollo de un sistema de programación es un proceso teórico que se divide en diversas etapas conceptuales; a saber: planificación, diseño, realización, ensayo y depuración. Durante la realización de un proyecto real de programación de un sistema, las distintas etapas no se suceden secuencialmente; por ejemplo, iniciado ya un proyecto, puede que cambien los planes y deban los programadores rediseñar parte del mismo.

La primera etapa importante en el desarrollo del soporte lógico es su planificación, donde se determina qué funciones han de realizarse y se acotan las diferentes situaciones a las que deben responder los programas del sistema. A medida que una tarea crece en complejidad, resulta más difícil especificar qué acción debe tomarse y cuándo. Un ejemplo simple nos ilustra este punto. Si se ha de contar el número de personas que hay en un pequeño auditorio, resulta intuitivamente obvio cómo ha de procederse, pero, si se trata de contar la multitud que llena un estadio, el método a utilizar requiere una mayor elaboración. Por ejemplo, ¿cómo se definen los límites del estadio? ¿Cómo se define una persona? (¿Debe contabilizarse el feto de una mujer embarazada?) Así, factores que raramente aparecen y que por consiguiente son irrelevantes en el primer caso pueden complicar el plan de un trabajo si crece su tamaño.

La especificación precisa de lo que debe hacer una defensa contra misiles balísticos es una tarea compleja. Por ejemplo, la instrucción "destrúyase todo misil soviético" bastaría si únicamente existieran misiles soviéticos y todos ellos tuvieran que destruirse bajo cualquier circunstancia. Pero el mundo no es tan sencillo. ¿Cómo pueden distinguirse los misiles soviéticos de los que no lo son? ¿Qué ocurre si un misil soviético se dirige hacia un objetivo de la República Democrática Alemana? Aunque las preguntas anteriores afectan sólo a un aspecto superficial del plan detallado de trabajo, exponen por primera vez un problema fundamental con el que se enfrentarán reiteradamente los técnicos: la dificultad de decidir si un aspecto particular de un programa es o no conveniente.

Además, ha de predecirse toda posible contingencia de forma exacta, y decidir cómo deben responder en cada caso los programas. Por ejemplo, qué pasos deben ejecutar los programas si los ordenadores de un escalón dado del sistema de dirección del combate fallan. ¿Cómo podría una defensa contra misiles balísticos distinguir el lanzamiento de una lanzadera espacial soviética del disparo de un ICBM? Dado el

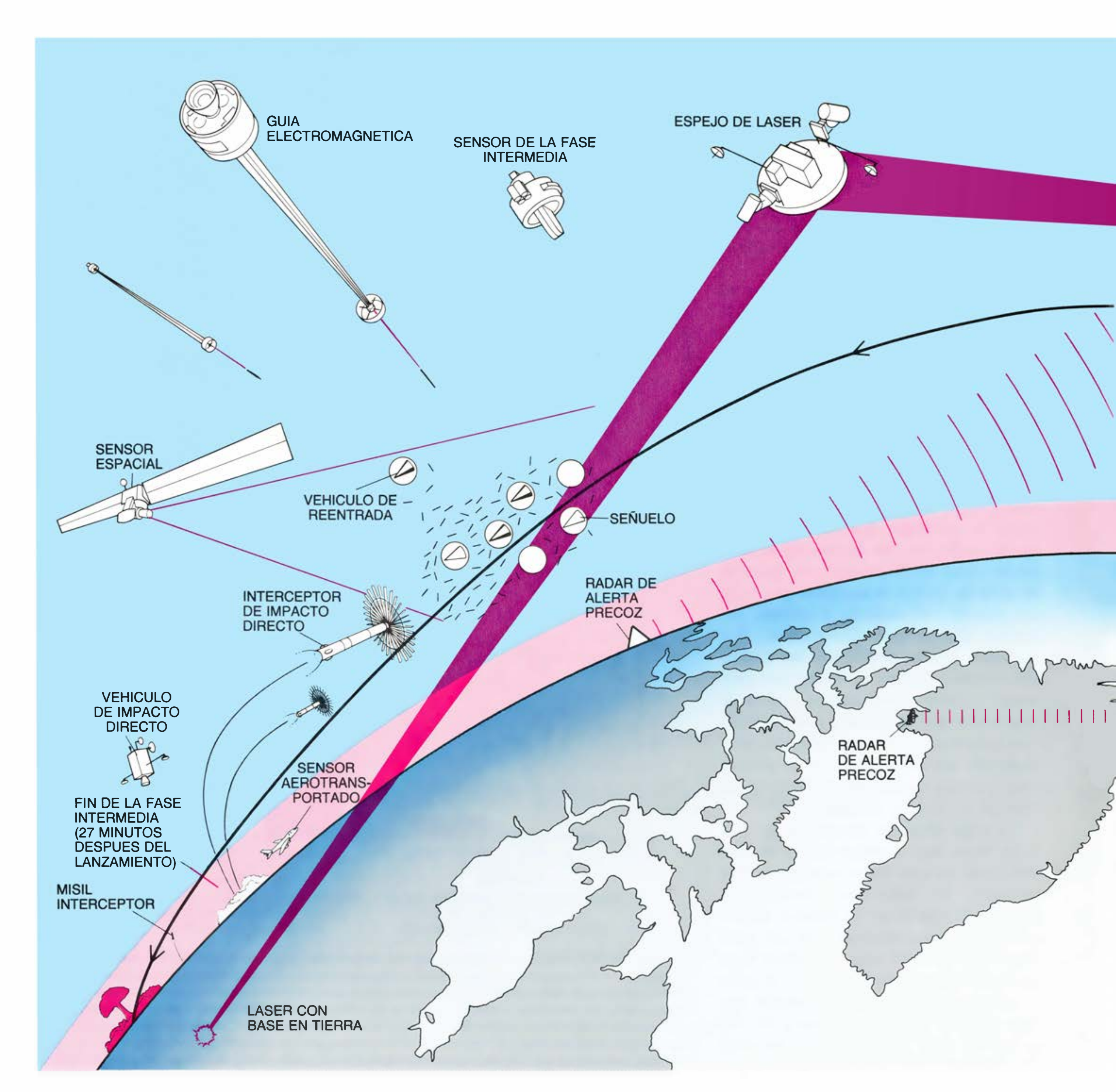


1. NIVEL DE ESFUERZO que requiere cada una de las cuatro etapas de desarrollo del sistema de programación para la defensa contra misiles balísticos (que según un estudio del Gobierno americano ocuparía 10 millones de líneas de instrucciones de programación); se muestra frente al trabajo necesario para producir los programas que controlan la unidad central de un ordenador (500.000 líneas). Las etapas no se suceden estrictamente de una forma secuencial, pues se producen solapamientos. Cuando nos referimos a la etapa final entedemos las pruebas y la eliminación de los errores previos a la entrega del producto; ya en uso el sistema se requerirían más pruebas con sus correspondientes eliminaciones de errores. En cada proyecto de desarrollo de programas, la columna de la izquierda representa estimaciones muy optimistas; la de la derecha representa puntos de vista menos optimistas. La función de escala que relaciona el nivel de esfuerzo con el tamaño del programa se ha tomado de *Software Engineering Economics*, de Barry W. Boehm.

número casi infinito de posibilidades, los responsables de la programación del sistema estarán sometidos a gran presión para que prevean todas las circunstancias posibles. La simple relación de todas las situaciones a las que podría enfrentarse un sistema BMD y de las acciones apropiadas que tuviera que tomar en cada caso ocuparían decenas de miles de páginas. A modo de comparación, el Reglamento de Impuestos de los Estados Unidos, que es la plasmación legislativa de la Ley Federal de Impuestos, ocupa alrededor de unas 3000 páginas. Así, la posibilidad de que en el plan detallado de los programas del BMD se consideren adecuadamente todas las situaciones potenciales es, como mínimo, tan improbable como la posibilidad de que no exista rendija legal alguna en las leyes de impuestos federales.

Para ilustrar este punto citaremos dos ejemplos. El 3 de junio de 1980 el Comando Norteamericano de Defensa Aeroespacial (NORAD) informó que los Estados Unidos eran atacados con misiles. El informe se atribuyó a un circuito defectuoso de un ordenador que generó señales incorrectas. Si al desarrollar los programas responsables de procesar estas señales se hubiera tenido en cuenta la posibilidad de un fallo en los circuitos, la falsa alerta se habría evitado.

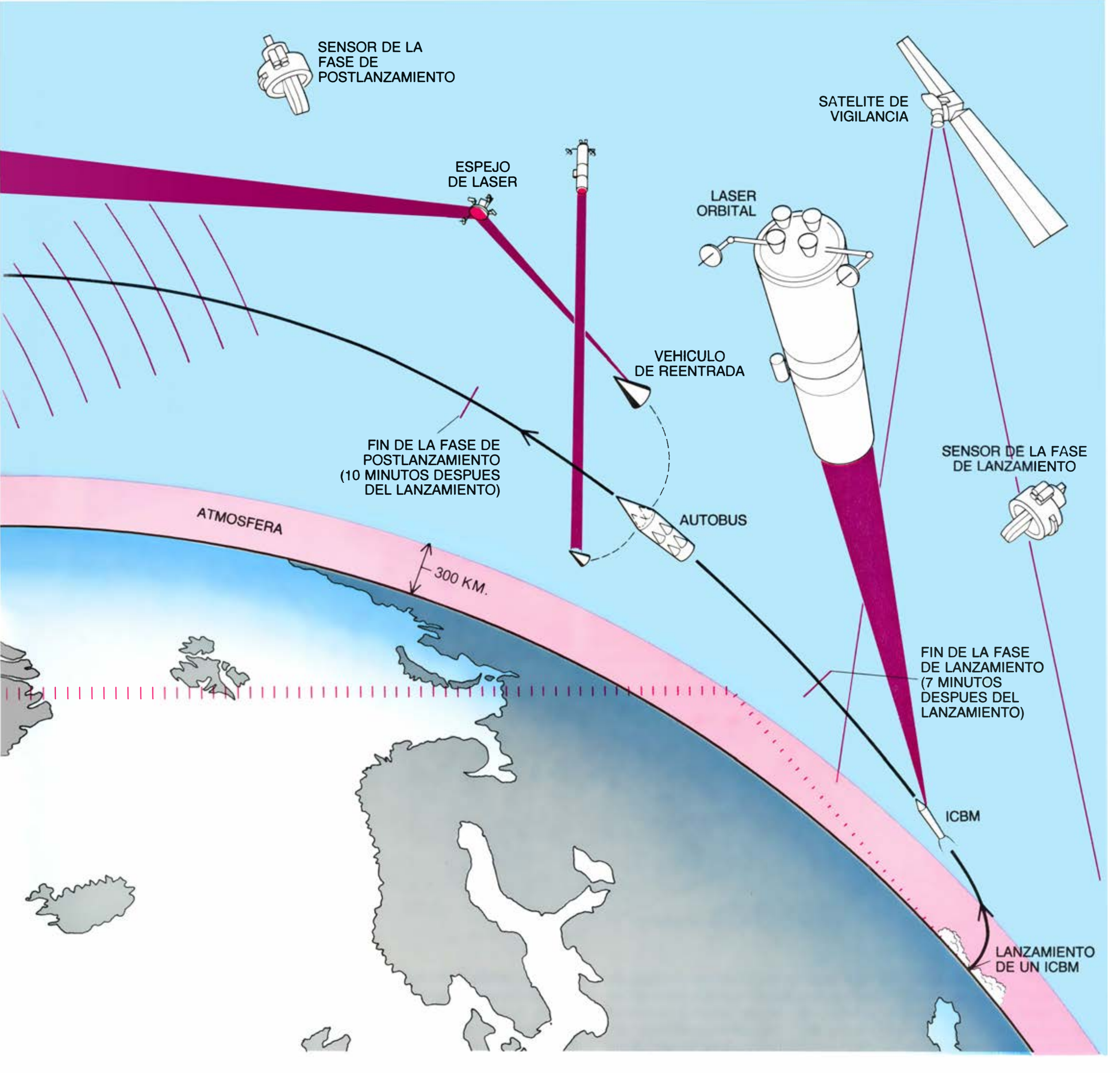
El segundo ejemplo, relacionado con



2. ENTRE LOS SOPORTES FISICOS PROPUESTOS para un sistema defensivo contra misiles balísticos se cuentan una gran variedad de armas interceptoras: láseres de gran potencia (con base terrestre o espacial), guías electromagnéticas y misiles. Además habría que contar también con sensores a bordo de satélites, aerotransportados y con base en tierra. Se atacaría a los misiles balísticos enemigos en cada una de las fases de su trayectoria: la fase de lanza-

el hundimiento del destructor británico H.M.S. *Sheffield*, salió a la luz pública tras los análisis que siguieron a la guerra de las Malvinas. De acuerdo con ellos, los sistemas de alerta del radar del buque se habían programado para que identificaran el misil Exocet como "misil amigo", puesto que el arsenal británico está dotado de ese tipo de armas. En consecuencia, el sistema ignoró las transmisiones del dispositivo de autoguía del Exocet hostil y permitió que éste alcanzara su objetivo, es decir, el *Sheffield*.

En la fase de diseño también son múltiples los posibles problemas. En ella debe resolverse la plasmación, en el ordenador, de los proyectos elaborados en la etapa de planificación; se desarrollan los algoritmos que han de conformar los programas, la secuencia con que han de discurrir ciertas acciones, etcétera. En su estructura teórica esta fase se asemeja a la que se enfrenta un arquitecto cuando, una vez determinadas las necesidades de una casa, ha de dibujar sus planos. Si el arquitecto se equivoca al evaluar la capacidad de drenaje del terreno, los cimientos quizá se inunden, provocando daños cuya reparación puede resultar muy costosa. De forma similar, también es caro corregir los errores de diseño de los programas si no se detectan en esta fase.

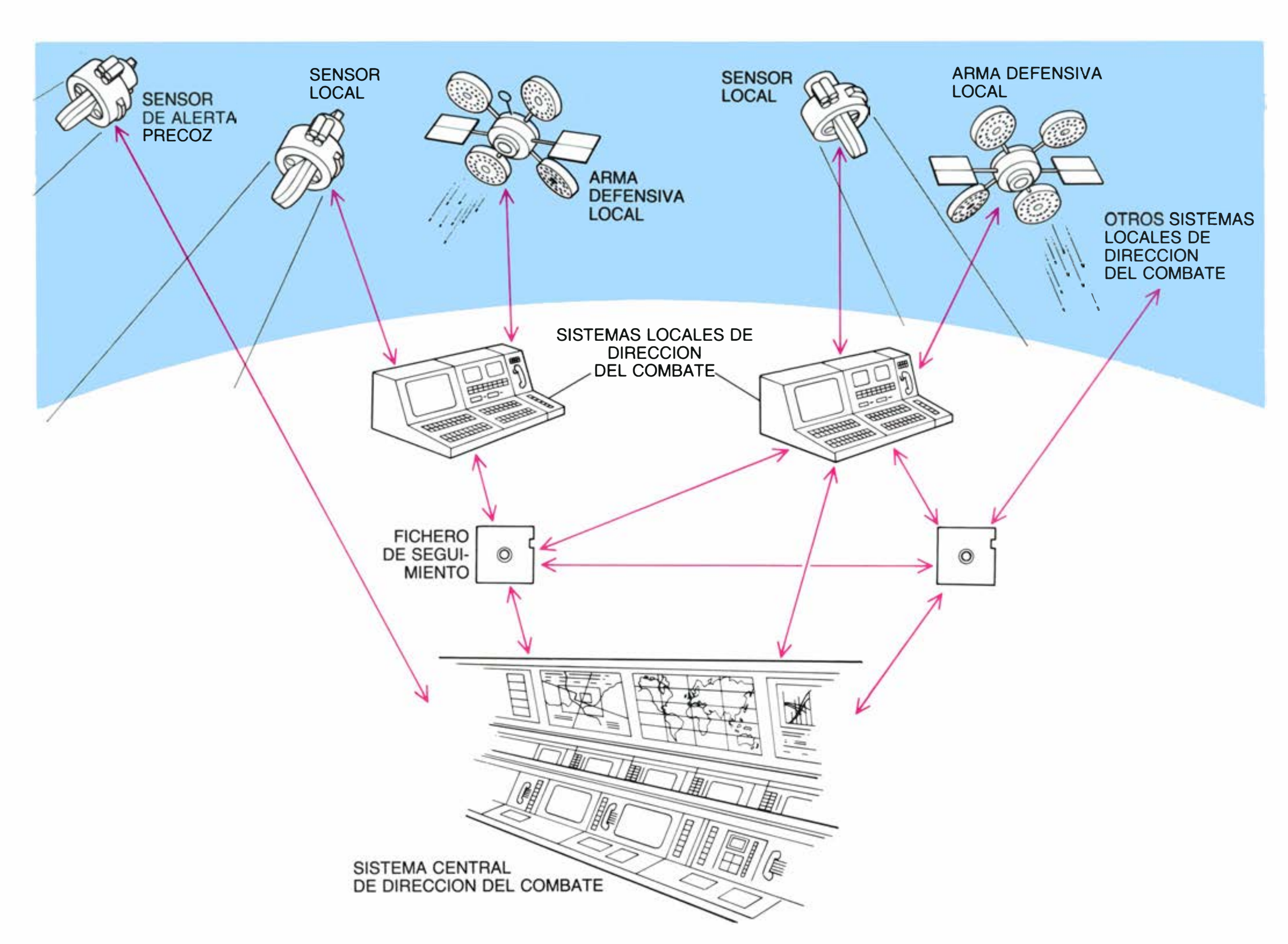


miento, durante la que el misil abandona la atmósfera en vuelo propulsado; la de postlanzamiento, en la que se van soltando los vehículos de reentrada, los señuelos y las piezas metálicas reflectoras del radar (*chaff*); la fase intermedia, durante la cual los diversos objetos viajan la mayor parte de su trayectoria en el espacio, y la fase terminal, en la que los vehículos de reentrada penetran en la atmósfera para detonar sobre los objetivos que se les haya asignado.

Ni siquiera importantes sistemas de programación se han visto exentos de tales errores de diseño. La cápsula espacial tripulada *Gemini V* erró su punto de amerizaje en 100 millas porque su programa de guía ignoró el movimiento de la Tierra respecto del Sol. En otro caso, se cerraron cinco reactores nucleares temporalmente porque un programa que evaluaba su resistencia a los terremotos aplicaba una suma aritmética de variables en lugar de la raíz cuadrada de la suma del cuadrado de las variables. No pueden, por tanto, predecirse ni la naturaleza ni la frecuencia de los errores de planificación y diseño; en realidad, esos errores sólo pueden eliminarse si los analistas los detectan con anticipación, tarea que resulta más ardua a medida que el tamaño o la complejidad del sistema aumentan. En consecuencia, detectar y corregir errores constituye uno de los aspectos fundamentales de todo proyecto de desarrollo de sistemas de programación. ¿Cómo detectarlos y cómo valorar la seguridad del funcionamiento de un sistema?

Dos técnicas han demostrado su eficacia en la valoración de la exactitud de un sistema de programación. Una de ellas es analítica y consiste en probar la corrección de los programas certificando matemáticamente que sus resultados se ajustan a ciertos criterios, especificados formalmente, al recibir datos de entrada conocidos. Con todo, aunque las pruebas de la correción de los programas ayudan a asegurar que el sistema de programación desarrolla sus especificaciones teóricas, no aseguran que éste cumpla su misión. Las pruebas completas son, como mínimo, comparables en tamaño con los programas que supuestamente verifican; los analistas se enfrentan entonces al problema de comprender una prueba tan complicada como el propio programa. Por otra parte, el programa de prueba no garantiza el tiempo que emplearía un sistema en realizar "cálculos a tiempo real" (factor crítico que se discutirá más adelante) ni revela la naturaleza de la respuesta del sistema al recibir datos imprevistos. Finalmente, las pruebas de corrección no aseguran que las propias especificaciones reales del programa sean correctas.

Una técnica más importante de valoración consiste en la prueba empírica. Obviamente, la defensa contra misiles balísticos no puede someterse a prueba empírica a gran escala en condiciones reales, ya que el coste de tal ensayo sería exorbitante. Más importante aún, la Unión Soviética no podría distinguir con seguridad entre el lanzamiento de gran número de misiles de



3. ENLACE entre sensores, armas y ordenadores; es decisivo en el sistema de dirección del combate de la defensa contra misiles balísticos (BMD). Los sensores y las armas constituyen un escalón de defensa que debería estar bajo el control de un computador local de dirección del combate. Tal sistema localizaría y seguiría el rastro de los objetivos potenciales, identificaría los blancos reales entre los señuelos y la chatarrería, asignaría sus armas a blancos específicos en tiempos específicos y se aseguraría de que el blanco se hubiera destruido. Toda la información sobre los objetivos se transmitiría a un "fichero de seguimiento" y al sistema contiguo de dirección del combate, así como a la dirección central de control del combate, que estaría en comunicación constante con las direcciones locales. Para cada una de ellas se especificaría bajo qué circunstancias precisas se ha de atacar un blanco determinado, se transmitiría la información del fichero de seguimiento y se coordinarían las acciones necesarias para proteger el propio sistema BMD. Además, el ordenador central de dirección del combate recibiría la alerta precoz concerniente al inicio del ataque. La red que se muestra podría romperse mediante un ataque al sistema central de dirección del combate.

ensayo y un ataque nuclear real por parte de los Estados Unidos. Por tanto, deben utilizarse formas más limitadas de evaluación empírica: pruebas a escala reducida, durante las cuales un sistema BMD actuase sólo contra unos pocos misiles, y pruebas de simulación, en las que un ordenador presentara amenazas a gran escala contra los objetivos que debe proteger un sistema BMD. Algunos expertos mantienen que tales pruebas hacen innecesarios los ejercicios reales a gran escala. Sin embargo, aunque tales ensayos incrementan la seguridad de un sistema, no aseguran el cumplimiento de los objetivos de la defensa total contra misiles balísticos.

Los problemas que surgen al integrar dos componentes individuales distintos en un sistema eficaz suelen aparecer sólo cuando se ensaya el sistema completo en situaciones límite; no se dan esas situaciones en ensayos a pequeña escala. El sistema WWMCCS (World Wide Military Command and Control System), red de comunicaciones utilizada por las autoridades civiles y militares para coordinar y transmitir información hacia y desde las fuerzas militares norteamericanas en campaña, constituye un buen ejemplo de lo mencionado anteriormente. En su utilización rutinaria, cuando la transmisión de mensajes es baja, el sistema WWMCCS funciona satisfactoriamente. Sin embargo, cuando el tráfico de mensajes es alto, la respuesta del sistema varía. Durante un ejercicio realizado en 1977, el WWMCCS se conectó a los sistemas de mando y control de varios centros regionales, actuando con una efectividad en la transmisión de mensajes de sólo el 38 por ciento.

Para superar las limitaciones de las pruebas de pequeña escala, a menudo se cuenta con simuladores que crean situaciones de ataque probables. Aunque la información obtenida de la simulación supera la conseguida mediante pruebas a pequeña escala, esta técnica, aplicada al ensayo de un sistema contra misiles balísticos, se ve limitada por varias razones. Primero, los simuladores no son capaces de reproducir las "signaturas", o parámetros de los fenómenos físicos asociados a diversos fenómenos, verbigracia, explosiones nucleares simultáneas, de una forma rápida y suficientemente exacta para poner a prueba la capacidad de respuesta del sistema defensivo. Por tanto, los analistas deben "confiar" en las signaturas de cada uno de los fenómenos que eligen para simular; es decir, los fenómenos se "preprocesan" antes de introducirlos en el sistema de defensa. Al emplear información preprocesada, no caben sorpresas durante la simulación que, por otra parte, debe probar cómo reacciona el sistema a circunstancias no previstas, disminuyendo, por tanto, el realismo de la prueba.

Aunque los avances en la tecnología de los computadores han incrementado enormemente su velocidad de cálculo, y por consiguiente es más fácil diseñar simuladores que modelen situaciones más realistas, el incremento de la velocidad de cálculo no ayudará a simular datos de fenómenos físicos de los que se dispone un conocimiento teórico y empírico inadecuado. Por ejemplo, atendiendo a los conocimientos actuales, los científicos no podrían modelar exactamente la explosión simultánea de varias ojivas nucleares en zonas cercanas bajo ciertas condiciones. Y más importante aún, los simuladores no pueden reproducir satisfactoriamente todos los ataques plausibles, ya que un enemigo inteligente elige sólo a última hora los parámetros de un ataque real. Por tanto, la confianza en un sistema BMD obtenida a partir de pruebas simuladas se apoya en la hipótesis de que los responsables de la simulación pueden predecir y reproducir electrónicamente la amplia gama de tácticas a las que podría recurrir el enemigo.

El desarrollo del sistema de defensa aérea Aegis ilustra las limitaciones de los ensayos basados en la simulación. El sistema de dirección de combate del Aegis se diseñó para rastrear cientos de objetos volantes en un radio de 300 kilómetros, asignando armas suficientes para destruir del orden de 20 de los objetivos situados dentro del alcance de sus misiles defensivos [véase "Armas inteligentes en la guerra naval", por Paul F. Walker; INVESTIGACIÓN Y CIENCIA, julio de 1983]. Se instaló el sistema en el crucero U.S.S. *Ticonderoga*. Una vez puesto en servicio el buque, el sistema se sometió a su primer ensayo real, durante el cual, por un error de programación que pruebas y simulaciones previas a pequeña escala no habían descubierto, fracasó en su intento de derribar 6 de 16 blancos. Dado lo limitado de aquellas pruebas, el sistema no se había enfrentado a más de tres blancos simultáneamente. En un ataque de proporciones próximas a los límites de diseño del sistema Aegis, posiblemente los resultados hubieran sido peores.

Son habituales los fallos durante las pruebas en condiciones reales; de hecho, durante el primer ensayo de un nuevo sistema es inevitable su funcionamiento defectuoso. En pruebas posteriores del Aegis, el sistema, ya reformado, no volvió a cometer los errores anteriores. Probablemente en futuras maniobras se descubran nuevos defectos, que a su vez se someterán a corrección, produciéndose en este proceso una mejora gradual del funcionamiento del sistema.

En contraste con el ejemplo anterior, la respuesta de una defensa total contra misiles balísticos no mejorará con la experiencia, ya que la prueba empírica a gran escala del sistema BMD consistiría en un ataque real a gran escala contra los Estados Unidos. Un sistema de dirección del combate más complejo que el Aegis, que ha de seguir más objetivos y que ha de operar en un tiempo mucho más reducido, es improbable que funcione mejor que el Aegis. Además, un error en plena batalla daría escasa opción, si alguna, a los responsables del sistema para tomar nota y aprender de la experiencia.

Con todo, los responsables del sistema de defensa contra misiles balísticos intentarán mejorar el funcionamiento del sistema, detectando y eliminando fallos, modificando sus características en respuesta a nuevas necesidades y añadiendo nuevos soportes físicos y programas al proyecto. Tales esfuerzos, junto con la dedicación de más personal al proyecto, constituyen lo que los especialistas denominan el mantenimiento del sistema de programación, que supone aproximadamente el 70 por ciento de los costes del ciclo de vida del desarrollo de un proyecto.

Todo sistema de defensa contra misiles balísticos presenta dos problemas de mantenimiento singulares. El primero y más importante es la eliminación de los fallos graves descubiertos después de que los programas estén ya en funcionamiento. El abanico de esos fallos puede extenderse desde un símbolo codificado incorrectamente hasta un defecto fundamental de diseño. El 19 de junio de 1985, el Organismo para la Iniciativa de Defensa Estratégica acometió un experimento sencillo: la tripulación de la lanzadera espacial debía situar su nave de forma que un espejo montado en uno de los lados de la nave reflejara un rayo láser emitido desde la cima de una montaña situada a 10.023 pies sobre el nivel del mar. El experimento falló porque el programa de ordenador que controlaba los mo-

vimientos del vehículo interpretó que la información sobre la localización del láser la recibía en millas náuticas, no en pies, y preparó la lanzadera espacial para recibir un rayo emitido desde 10.023 millas náuticas sobre el nivel del mar, en una inexistente montaña. El pequeño error de procedimiento tuvo poca trascendencia para la prueba, pues un segundo intento realizado pocos días después terminó con éxito, pero muestra cómo hasta los más simples errores pueden dar al traste con una misión.

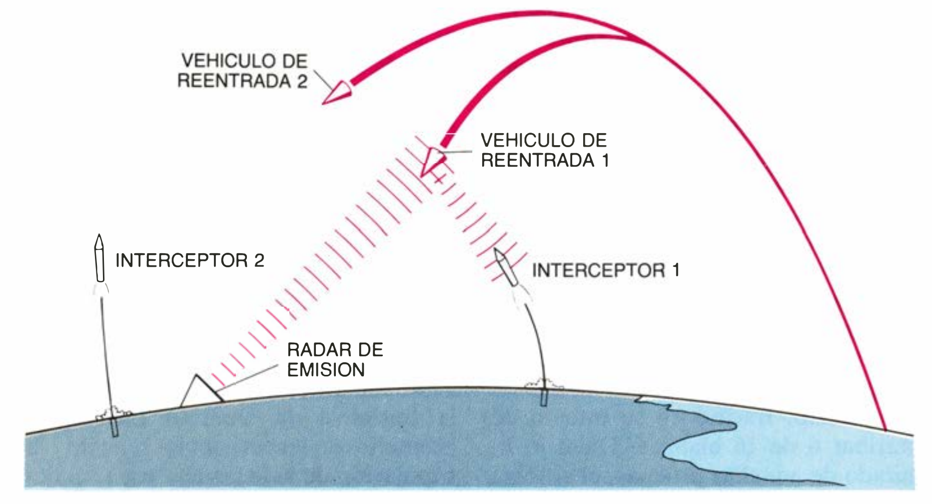
Aunque en el caso anterior pudo rectificarse el fallo fácilmente, depurar un sistema de defensa contra misiles balísticos requerirá un esfuerzo considerablemente mayor. En particular, durante un ataque, el sistema de programas de la dirección del combate debe recibir y procesar información de tal forma que actúe al mismo ritmo que se suceden las circunstancias externas al ordenador. Es decir, los programas deben trabajar a "tiempo real" [*véase la figura 4*]. Uno de los problemas a considerar cuando se trata de eliminar fallos en un paquete de programas a tiempo real es el de determinar por qué a veces unos programas que, con una configuración o equipamiento determinado funcionan bien, con otra configuración, ligeramente distinta de la anterior, no lo hacen. Por ejemplo, cierto misil que puede lanzar sin dificultad a velocidades supersónicas el avión F-4G Wild Weasel no funciona con el F/A-18 Hornet: un sistema de programación que funciona en el modelo F-4G no es enteramente compatible con la aviónica instalada en el modelo F/A-18.

Un segundo problema de especial significado en la eliminación de fallos en los programas a tiempo real es que a los analistas a menudo les resulta difícil producir recurrencias de errores, práctica esencial para localizarlos. La respuesta a tiempo real a menudo está determinada por factores como el tiempo de llegada de la información de los sensores, y esos factores determinantes del comportamiento de un programa no siempre pueden reproducirse con la suficiente exactitud para encontrar y eliminar el fallo. De hecho, para hallar un error hace falta conocer las circunstancias precisas que lo producen y, dado que ello no es siempre posible en grandes sistemas a tiempo real, los errores de programación se identifican en la práctica sólo en un sentido probabilístico. Los grandes programas de ordenador se desarrollan por medio de una serie de cambios conocidos sólo de forma incompleta. A partir de cierto punto, el programador no puede predecir con seguridad las respuestas del sistema y no le queda más remedio que confiar en que se alcance el resultado deseado.

Existe una complicación final en la eliminación de fallos de los programas a tiempo real: aun habiéndose localizado un error, puede que los intentos de eliminarlo no fructifiquen, pues la probabilidad de que al eliminar un error se introduzca otro (o más de uno) oscila entre el 15 y el 50 por ciento. Además, la mayoría de los errores de diseño que se revelan una vez puesto en servicio el sistema lo hacen tras un uso frecuente. La experiencia obtenida en el trabajo con grandes programas de control (que constan de entre 100.000 y dos millones de líneas de instrucción) sugiere que la probabilidad de introducir un error importante al corregir otros es tan grande que sólo debe enmendarse una pequeña fracción de los fallos originales.

En el contexto de la defensa total contra misiles balísticos debemos reflexionar acerca de las consecuencias de un fallo que se manifestara de una forma impredictible e infrecuente. Los detalles del primer intento de lanzamiento de la lanzadera espacial constituyen un buen ejemplo. La lanzadera, cuyos programas operativos a tiempo real ocupan unas 500.000 líneas de instrucciones, fracasó en su primer intento de despegue por un problema de sincronización entre los ordenadores de control de vuelo. El error de programación causante del fallo se introdujo al corregir otro error previo dos años antes, y se hubiera manifestado, por término medio, en una de cada 67 pruebas.

Además de detectar y eliminar errores, un segundo aspecto del mantenimiento de un sistema de programación constituye también un obstáculo al desarrollo del sistema defensivo. El Equipo de Estudio de Tecnologías Defensivas (DTST), organismo auspiciado por el Departamento de Defensa para examinar la fiabilidad de una defensa total contra misiles balísticos, sostiene que el sistema defensivo completo necesitará un mínimo de 10 millones de líneas de instrucciones de programación. Comparándolo con el sistema Aegis, éste resulta un orden de magnitud más pequeño. Si la estimación del DTST está infravalorada en un factor de dos, hasta el proyecto de desarrollo de programas más optimista comportaría más de 30.000 años-hombre de trabajo o, lo que es lo mismo, el trabajo de 3000 programadores y analistas como mínimo, durante diez años. Por esa razón, el proyecto ha de prever un movimiento de personal que reducirá la memoria institucional del mismo. Así, durante los cambios de personal podría omitirse algún detalle esencial, como el de actualizar algún subprograma. Un ejemplo trágico de ese tipo de errores de coordinación ocurrió en 1979, cuando un avión de Air New Zealand chocó contra una montaña de la An-



4. UN PROGRAMA A "TIEMPO REAL" ha de ejecutarse al ritmo que define una escala de tiempos externa al soporte físico del ordenador. Un computador que controle el radar que emite señales hacia un vehículo de reentrada, para que un misil interceptor capte la señal reflejada, tendría que mantener la señal de radar sobre el vehículo de reentrada hasta que éste hubiera sido destruido. Si el radar se desviara de su objetivo demasiado pronto, el interceptor podría "perder" su objetivo, en especial si éste estuviera preparado para evadir interceptores. Probablemente el ordenador habría de apuntar el radar hacia otros vehículos, para interceptarlos. El programa habría de desviar el radar hacia un segundo vehículo una vez destruido el primero, y ello con la rapidez necesaria para que el segundo interceptor alcance a derribar su objetivo.

tártida porque su tripulación no fue advertida de que los datos de entrada que señalaban el plan de vuelo del ordenador de navegación se habían cambiado.

Otra posibilidad es que agentes soviéticos trataran de sabotear el proyecto. Podrían introducir deliberadamente en el sistema defectos difíciles de detectar, que aparecieran sólo durante un ataque real pero no durante las pruebas. Tampoco se puede dejar de lado la posibilidad de que un programa pueda contener algún tipo de "bomba retardada" que produjera un fallo de los programas en un momento crítico. La precaución obvia contra tales acciones consistiría en imponer estrictas medidas de seguridad que, por añadidura, inhibirían la comunicación entre el personal asignado a diferentes partes del sistema, incrementándose, por tanto, la posibilidad de importantes errores de programación no intencionados.

Fortaleciendo la seguridad no se impide que la Unión Soviética obstaculice el sistema mediante otros medios. No existe razón para creer que los funcionarios soviéticos que se enfrentaran a un sistema BMD norteamericano renunciaran a disimular las características de sus misiles balísticos y de sus cabezas nucleares o señuelos durante las pruebas o durante su utilización real. Sin datos veraces sobre los misiles soviéticos, no cabría seguridad alguna acerca del funcionamiento del sistema de defensa americano. Es más, los rusos podrían desarrollar tácticas nuevas o armas que forzaran a los analistas norteamericanos a reprogramar el sistema de dirección del combate para afrontar nuevas amenazas. La acumulación de varios cambios a lo largo del tiempo probablemente conduciría a muchas interacciones imposibles de predecir y, en consecuencia, a la necesidad de un nuevo diseño del sistema.



5. MISIL SM-2 que forma parte del sistema de defensa aérea Aegis de la Armada de los Estados Unidos. El sistema de dirección de combate del Aegis se diseñó para que siguiera cientos de objetivos aéreos y realizara hasta 20 intercepciones simultáneas. Debido a defectos de programación, durante su primer ensayo en condiciones reales falló en el derribo de 6 de 16 objetivos. Posteriormente se han reparado los errores del sistema. Una defensa total contra misiles balísticos, que constituye un sistema mucho más complicado y ambicioso, debería comportarse mucho mejor que el sistema Aegis, ya que su funcionamiento habría de ser correcto al primer intento; un fallo importante en la programación del sistema durante la primera prueba global posiblemente no daría oportunidad a nuevas rectificaciones. Dada la imposibilidad de efectuar ensayos a gran escala, el primer ensayo de un sistema BMD sería un ataque real a gran escala de misiles balísticos.

Se ha sugerido también que la elaboración de programas para sistemas extremadamente complejos podría aligerarse mediante el concurso de sistemas inteligentes y de programación automática. En este sentido, un informe de la Agencia para Proyectos de investigación Avanzados de la Defensa, DARPA, apuntaba la posibilidad de aplicar sistemas inteligentes a la defensa contra misiles balísticos. Los sistemas inteligentes son descripciones detalladas, expresadas como reglas de programación, de los diversos algoritmos deductivos que utilizan los expertos humanos para planear o tomar decisiones en cada una de sus especialidades. Así, los sistemas inteligentes utilizan algoritmos informales o reglas rutinarias. Por ejemplo, uno de esos sistemas que tuviera la siguiente frase como dato: "El lanzamiento de un ICBM soviético supone una amenaza para los Estados Unidos", podría invertir el orden de la frase, y utilizarlo como regla legítima de deducción: "Si existe una amenaza contra los Estados Unidos se debe al lanzamiento de un ICBM por parte soviética". La validez de esta herramienta deductiva no puede aprobarse o desecharse usando el razonamiento habitual de la lógica formal; la frase no es cierta siempre, pero es prudente la mayoría de las veces. La fiabilidad de los procedimientos de razonamiento informal de los sistemas inteligentes de programación deja abierta la posibilidad de que se produzca un error debido a contradicciones conceptuales profundamente arraigadas en el sistema.

Hasta la fecha, la investigación en sistemas expertos se ha localizado en áreas bien definidas, como la bioquímica y la medicina interna. En esas áreas, la pericia humana se ha desarrollado lo suficiente para que esos sistemas resulten satisfactorios. Lo que no obsta para que esas técnicas tengan un alcance limitado. Quienes recomiendan aplicar sistemas inteligentes a la dirección del combate en el sistema de defensa contra misiles balísticos ignoran que la pericia humana es fruto de la experiencia, y nadie domina los ataques de misiles nucleares por su experiencia práctica.

Igualmente improbable es la idea de que la programación automática (el uso de programas de ordenador para escribir otros programas) pueda constituir una solución del problema. El comandante de las Fuerzas Aéreas Simon Worden, ayudante especial del director del Organismo para la Iniciativa de Defensa Estratégica, acaba de manifestar: "Vamos a desarrollar nuevos sistemas de inteligencia artificial, para escribir los programas del sistema. La eliminación de sus errores también será cometido del sistema de inteligencia artificial". Frases como esa inducen a error. La función fundamental de la programación automática consiste en aliviar las dificultades técnicas que

comporta traducir en programa las especificaciones del diseño del sistema y modificar una codificación ya existente. Sin embargo, aproximadamente la mitad de todos los errores de programación proceden de elecciones humanas y de decisiones tomadas durante las fases de planificación y diseño, que no pueden delegarse en una programación automática.

Basándonos en el análisis anterior, mostramos ahora dos situaciones que ilustran cómo podrían fallar los programas de un sistema de defensa total contra misiles balísticos. Supongamos que una estación de combate que ha interceptado con éxito dos misiles durante un ejercicio operativo utilizando una guía electromagnética se enfrenta ahora a un ataque de misiles soviéticos a gran escala. Inicialmente, los proyectiles lanzados desde la estación de combate destruyen sus objetivos pero, desafortunadamente, en la fase de diseño del sistema no se tuvo en cuenta que cuanto menor es la masa de un cuerpo mayor es su retroceso. A medida que se van lanzando proyectiles, el retroceso es cada vez mayor, lo cual afecta a los algoritmos de puntería de la guía electromagnética. Así, los últimos proyectiles son demasiado lentos y no alcanzan sus objetivos en el momento adecuado. Supongamos, también, que durante la fase inicial de una guerra nuclear, los dirigentes de las dos superpotencias acuerdan un alto el fuego. Los americanos descubren que el capitán de un submarino no recibirá el mensaje del cese de hostilidades a tiempo, pero confían en que los satélites de defensa estadounidenses derriben todos los misiles lanzados erróneamente. Cuando el submarino ha lanzado sus misiles, se advierte trágicamente que no se había considerado la posibilidad de que los satélites destruyeran armamento propio. Las armas explotan en la Unión Soviética con consecuencias catastróficas.

Es poco probable que llegue a producirse alguno de los fallos descritos en este artículo, pues ya son errores predecibles. El problema fundamental no consiste en saber si el sistema tiene un determinado error, sino en conocer con qué probabilidad puede contener uno cualquiera de los millones de errores potenciales. La principal inquietud se centra, pues, en aquellos fallos potenciales cuya predicción sigue siendo imposible.

Todos los ejemplos citados aquí guardan relación con programas planeados, diseñados, redactados y que sufrieron un proceso de depuración por experimentados ingenieros de programación: la mayoría, si no todos, hubieron de corregirse. La técnica general para subsanar estos errores, su descubrimiento mediante el uso real del programa y su posterior correción, difícilmente será útil en el desarrollo de un sistema BMD. Así, un sistema de defensa total contra misiles balísticos requiere no sólo que los programas funcionen bien la primera vez en un entorno impredecible y sin haberse sometido a pruebas de gran escala, sino también que sus diseñadores estén seguros de que lo haga. La administración Reagan ha manifestado que las pruebas empíricas son esenciales para mantener la seguridad de las armas nucleares y, por ese motivo, se ha opuesto a una prohibición global de las pruebas. Cabe preguntarse quién confiará en un sistema de defensa total contra misiles balísticos, que, cuando menos, será tan complicado como lo son ahora las armas nucleares, sin que haya demostrado su capacidad de superar pruebas comparables.

Los partidarios de la Iniciativa de Defensa Estratégica aducen que, incluso sin pruebas empíricas a gran escala, es deseable un sistema de defensa contra misiles balísticos, porque la posibilidad de que el sistema funcionara razonablemente bien disuadiría a los dirigentes soviéticos a plantearse la posibilidad de un ataque. Otros argumentan que los objetivos de la Iniciativa de Defensa Estratégica son realmente más modestos que los de un sistema de defensa total contra misiles balísticos, y que quizás un sistema BMD resulte más apropiado para proteger asentamientos militares, como los silos de misiles. Naturalmente, frente a objetivos más limitados que los de la defensa total contra misiles balísticos, la perfección ya no es un requisito imprescindible y, por tanto, la ejecución de ensayos resulta mucho más fácil. Aunque un programa con objetivos más limitados tenga una probabilidad más alta de éxito, la propia limitación plantea un nuevo problema, en concreto, la contradicción con la visión planteada por el presidente Reagan de eliminar la amenaza de los misiles balísticos nucleares. En consecuencia, nos parece oportuno juzgar la fiabilidad de la Iniciativa de Defensa Estratégica en razón del reto que el presidente plantea a la comunidad científica. Y no parece previsible ninguna tecnología de programación que, hoy por hoy, dé soporte a un sistema de defensa contra misiles balísticos.

Conjugación de fase óptica

En la vida cotidiana el tiempo transcurre siempre hacia adelante. Sin embargo, la situación es cualitativamente diferente en el caso del movimiento ondulatorio: las trayectorias de las ondas luminosas pueden ser "invertidas temporalmente"

Vladimir V. Shkunov y Boris Ya. Zel'dovich

Imagínese una nadadora en un trampolín, preparándose para dar un salto. Una carrera rápida, un vuelo..., pero debido a un pequeño error técnico su cuerpo penetra en el agua con un ángulo incorrecto, dándose un gran chapuzón y originando ondas que avanzan a partir del punto de contacto con el agua. ¡Qué maravilloso sería poder invertir el tiempo, con el fin de corregir el error y obtener una mayor puntuación! El agua proyectada volvería a unirse, las ondas regresarían al punto de contacto, la saltadora sería arrojada fuera del agua y ascendería hasta el trampolín y la superficie quedaría tan lisa como lo estaba antes del salto. Desgraciadamente, aunque este supuesto se puede obtener fácilmente con la ayuda de un proyector cinematográfico, el proceso de inversión temporal que representa contradice nuestra experiencia cotidiana. Existe una buena razón para ello: la secuencia de acontecimientos que acabamos de describir viola el segundo principio de la termodinámica (la ley que establece que los sistemas tienden a la máxima entropía).

Sin embargo, dicha secuencia se puede realizar con éxito si lo que se considera es el movimiento de la luz o cualquier otra radiación electromagnética. Este fenómeno es posible gracias a una importante propiedad de los rayos luminosos, conocida desde hace mucho tiempo: el carácter reversible de su propagación. Para cada haz luminoso con una estructura arbitraria de rayos existe un posible haz con "tiempo invertido" cuyos rayos siguen las mismas trayectorias, aunque en sentido opuesto, de forma análoga al movimiento de una película hacia atrás. El éxito de la inversión del movimiento ondulatorio débese a la gran simplificación del problema: los movimientos térmicos y mecánico-cuánticos de los átomos y electrones que radian y refractan la luz no necesitan ser invertidos. En la práctica, basta con invertir el comportamiento temporal de los parámetros macroscópicos que describen el movimiento medio de un gran número de partículas.

La existencia de haces invertidos tiene consecuencias muy importantes. Es evidente, por ejemplo, que un haz ideal, es decir, libre de distorsión y con una divergencia mínima, se puede degradar cuando se transmite a través de materiales no homogéneos (tales como una lámina de vidrio de espesor no uniforme). La propiedad de reversibilidad implica que es posible crear un haz "antidistorsionado" que se hace ideal después de ser transmitido hacia atrás a través de dichos materiales no homogéneos. La técnica por la que los haces en cuestión se crean y se manipulan se conoce con el nombre de conjugación de fase óptica. Las ondas que constituyen dicho haz se denominan ondas de fase conjugada.

Antes de describir las propiedades de una onda de fase conjugada vamos a exponer algunas nociones básicas del movimiento ondulatorio. A medida que un conjunto de ondas se mueve en el espacio, sus oscilaciones llegan a los diferentes puntos del espacio en tiempos diferentes. De los puntos en que las oscilaciones son sincrónicas se dice que están en fase. La fase representa el valor del período en relación con cierta posición inicial. Las superficies que unen a los puntos con una misma fase se conocen como frentes de onda. Una propiedad importante de los frentes de onda consiste en que dichas superficies son perpendiculares a la dirección de propagación. Los frentes de onda de las ondas planas son planos, y esferas concéntricas, los frentes de onda de las ondas esféricas. Los frentes de onda de los haces luminosos reales pueden adquirir formas y topologías bastante complicadas.

El concepto de frente de onda se puede utilizar para comprender las propiedades de una onda de fase conjugada. Supongamos que hacemos una fotografía de una onda luminosa en la que el haz se propaga de izquierda a derecha [*véase la figura 3*]. En virtud de la reversibilidad de la propagación de la onda, al examinar esta fotografía no seríamos capaces de decir si el sentido de propagación es de izquierda a derecha o de derecha a izquierda. Sin embargo, si el haz se propagara de derecha a izquierda (es decir, si el haz se hubiera invertido), los frentes de onda se habrían invertido respecto a la dirección del haz. Este es el motivo por el cual el proceso de generar una onda invertida se conoce también con el nombre de inversión de los frentes de onda.

La relación entre los frentes de onda de dos ondas mutuamente invertidas es análoga a la relación entre las posiciones de dos ejércitos enemigos sobre un mapa militar. El frente de cada ejército coincide con el del otro, y las direcciones de los movimientos deseables son opuestas. Podemos decir que las líneas del frente están invertidas entre sí: una parte convexa del frente de uno de los ejércitos corresponde a una parte cóncava del otro.

Utilizando un lenguaje diferente, diremos que la diferencia de fase entre dos puntos cualesquiera del haz invertido tiene un signo opuesto al de la diferencia de fase entre esos mismos puntos del haz original. La operación matemática de cambiar el signo de la fase recibe el nombre de conjugación. Por esta razón se acuñó el término "conjugación de fase óptica" y se introdujo posteriormente en la literatura científica.

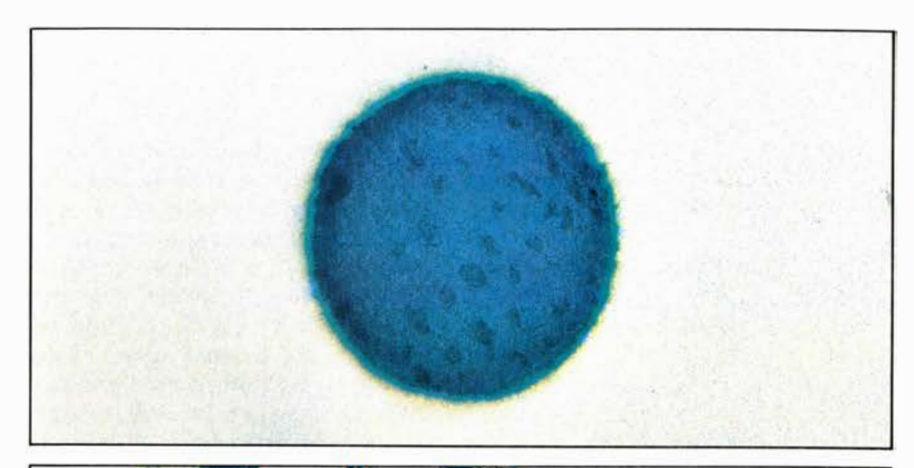
¿Cómo se conjuga una onda? (Es decir, ¿cómo se invierte?) Obtener la conjugación de fase de una onda plana es fácil: la utilización de un espejo plano da lugar a la reflexión de la onda hacia atrás. La conjugación de una onda esférica no resulta mucho más difícil. En este caso, se monta un espejo cóncavo con forma de sección

esférica de manera que el centro del espejo corresponda a la fuente de la onda. En cada punto del espejo los rayos inciden perpendicularmente y, por tanto, son reflejados exactamente hacia atrás.

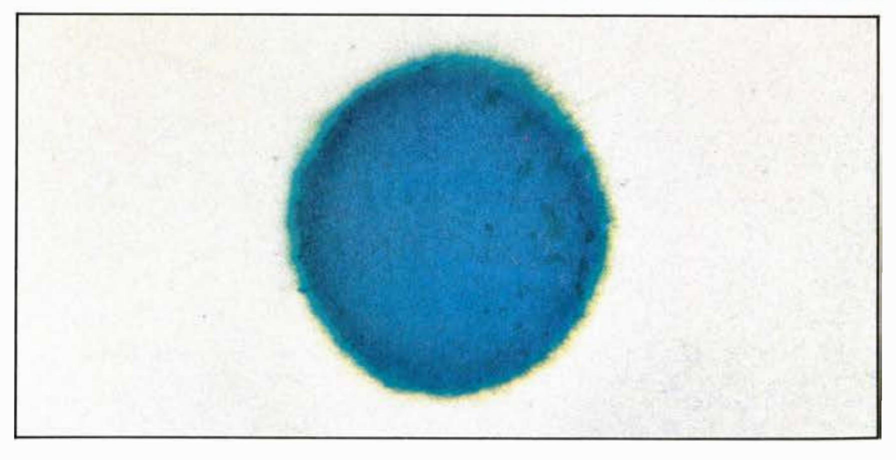
Para conjugar un haz con un frente de onda arbitrario se podría utilizar en principio un espejo cuyo perfil coincidiera con el frente de onda. Desgraciadamente resulta difícil poner en práctica este método. En primer lugar, sería necesario construir un espejo diferente para cada haz incidente. En segundo lugar, la forma del frente de onda de un haz de láser puede cambiar durante un impulso breve; por consiguiente, la forma del espejo debería cambiarse de manera continua para que en todo momento se ajustara a la forma de la onda. Finalmente, la precisión requerida para construir y colocar tal espejo sería extremadamente alta.

Para producir una onda de fase conjugada se requiere la utilización de un medio o una superficie cuyas propiedades resulten afectadas por las características de las ondas que inciden en él. Esta dependencia permite que dicho medio o superficie se ajuste automáticamente al haz incidente con tal exactitud que, bajo ciertas condiciones, se origine un haz de fase conjugada. Afortunadamente tales materiales existen y se conocen con el nombre de materiales ópticos no lineales.

La difusión de Brillouin estimulada y la mezcla de cuatro ondas son dos métodos basados en el empleo de estos materiales y ampliamente utilizados para la conjugación de fase. La difusión de Brillouin estimulada, uno de los efectos más bellos de la óptica no lineal, fue descubierta en 1964 por Raymond Y. Chiao, Boris P. Stoicheff y Charles H. Townes, mientras trabajaban en el Instituto de Tecnología de Massachusetts. El efecto tiene lugar cuando se dirige un haz luminoso sobre un medio transparente tal como un cristal, vidrio, líquido o gas comprimido. La luz de baja intensidad atraviesa estos medios sin atenuación apreciable. Sin embargo, el comportamiento de un haz luminoso de alta in-







1. HAZ LUMINOSO con la fase conjugada, es decir, con la fase "invertida temporalmente"; puede compensar las distorsiones originadas por un medio ópticamente no homogéneo o irregular (un vidrio deslustrado). Un haz de láser altamente direccional (*arriba*) resulta distorsionado y se convierte en un haz degradado (*centro*). A continuación, el haz degradado se invirtió por conjugación de fase óptica. La transmisión hacia atrás del haz con la fase conjugada a través del medio irregular o no homogéneo restablece la calidad del haz original (*abajo*).

tensidad resulta sorprendente. Cuando se alcanza una potencia umbral del orden del millón de watts, el haz se refleja hacia atrás casi por completo. Aunque dicha potencia es bastante elevada, se puede obtener con facilidad utilizando un láser pulsante de laboratorio.

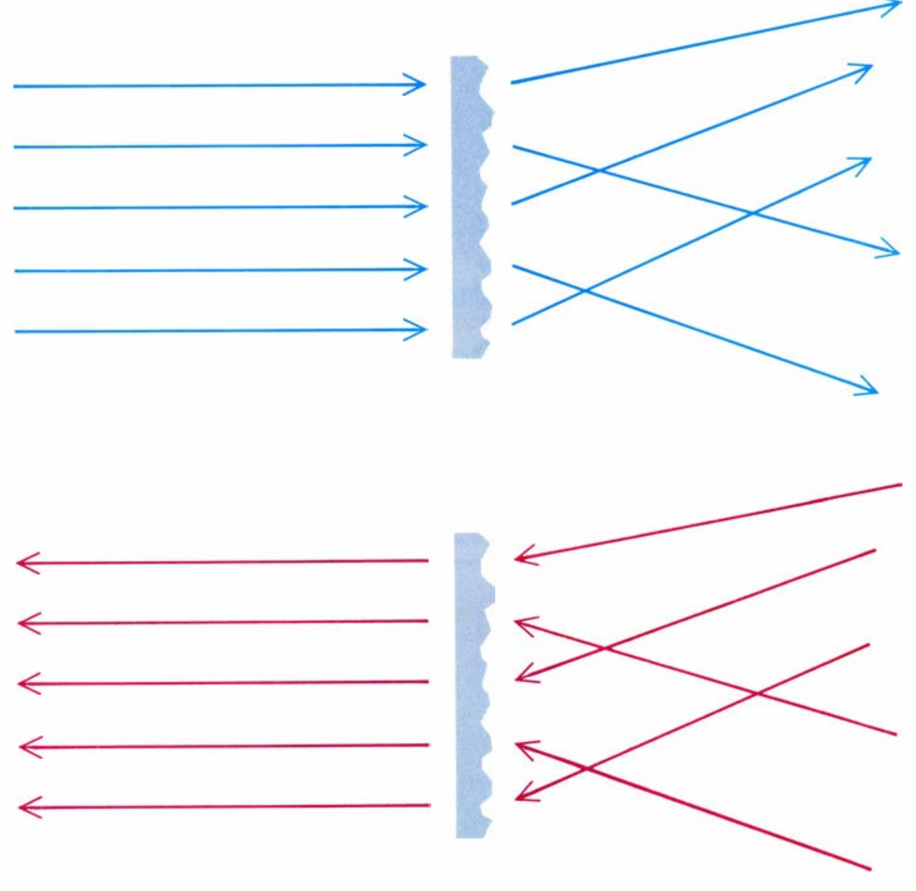
El haz reflejado es la consecuencia del proceso que da lugar a la difusión de Brillouin (denominada así en memoria del físico francés Louis Marcel Brillouin). La difusión de Brillouin se produce cuando una onda sonora incide sobre un sólido, líquido o gas. La onda sonora origina ciertas alteraciones periódicas en la densidad del material en que se propaga. Como resultado, se forman una serie de zonas de compresión y enrarecimiento, alternativamente, que se mueven con la onda a través del material.

Dado que las zonas de compresión son más densas que las zonas de enrarecimiento, el comportamiento de una onda luminosa que incida en el material es diferente en estas dos zonas. Concretamente, el índice de refracción del primer tipo de zonas es ligeramente diferente del índice de refracción del segundo. (El índice de refracción de un material es el cociente entre la velocidad de la luz en el vacío y la velocidad de la luz en el material.) Si la separación entre las zonas es exactamente igual a la semilongitud de onda de la luz incidente, la luz se reflejará. Este tipo de reflexión resultará familiar para los que hayan observado una capa delgada de aceite sobre agua y advertido su superficie coloreada con todos los colores del arco iris. En cada punto de esta capa se refleja un color mejor que los otros: aquel cuya semilongitud de onda equivale al espesor de la capa en cuestión en dicho lugar. Como el espesor de la capa de aceite varía de un punto a otro, se reflejan en cada punto diferentes colores.

En la difusión de Brillouin estimulada, la onda sonora, o variación de la presión y la densidad, no se aplica externamente al material, sino que es estimulada internamente por pares de ondas luminosas propagándose en sentido contrario. De la misma forma que el sonido consiste en una onda de presión y densidad, la luz consiste en un campo eléctrico en movimiento. Un campo eléctrico puede comprimir el material; este fenómeno se conoce con el nombre de electrostricción. En consecuencia, cuando un campo eléctrico se mueve a través de un material con la velocidad del sonido, puede dar lugar a una onda sonora. Dicho campo eléctrico puede generarse por interferencia de dos haces ópticos que avanzan en direcciones opuestas si la diferencia de frecuencias entre los dos haces es igual a la frecuencia del sonido. En el caso de una difusión de Brillouin estimulada uno de estos haces es el haz luminoso incidente.

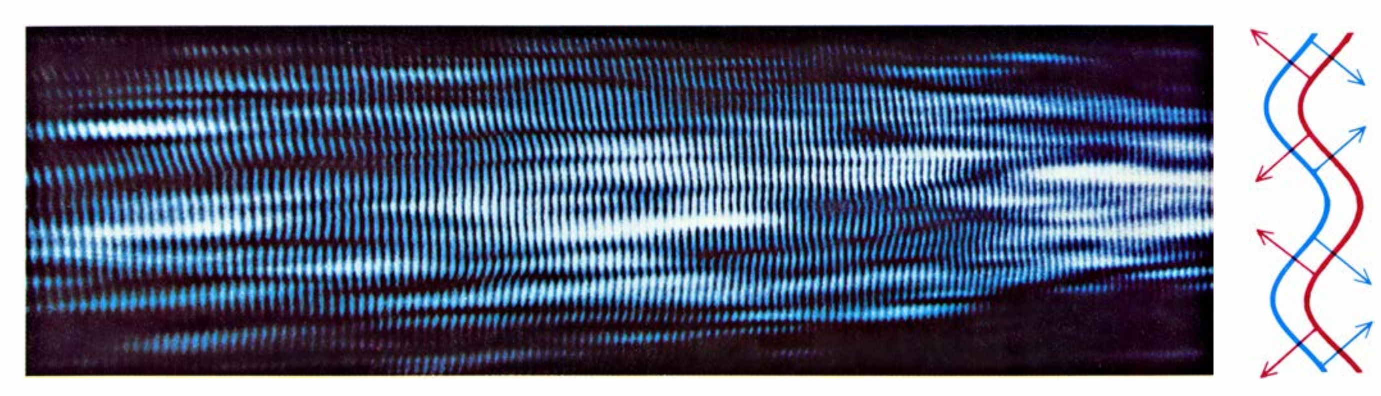
El otro haz luminoso procede de la difusión del haz incidente por las pequeñas fluctuaciones de densidad, estocásticamente distribuidas en el medio (es decir, ondas sonoras fluctuantes térmicamente). Cuando la frecuencia y la dirección de una onda difundida son las adecuadas, la onda interfiere con el haz incidente y amplifica las variaciones de presión y densidad en el material. Estas variaciones conducen subsiguientemente a la reflexión de una pequeña parte del haz incidente. La fracción reflejada se interfiere a su vez con el haz incidente, dando lugar a nuevas variaciones de presión y densidad. Estas variaciones conducen a más reflexiones del haz incidente. Las reflexiones aumentan exponencialmente con la distancia, hasta que un haz reflejado emerge del material. Sin embargo, dado que la amplificación depende de la intensidad del haz incidente, un requisito previo para obtener dicho haz reflejado consiste en que la potencia del haz incidente supere cierto valor umbral.

La conjugación de fase por difusión de Brillouin estimulada fue obtenida por vez primera en 1972 por Valery V. Ragul'skii, Vladimir I. Popovichev, Fuad S. Faizullov y uno de nosotros (Zel'dovich) en el Instituto de Física P. N. Lebedev de Moscú. El factor principal en la consecución de la conjugación de fase consistió en el uso de una placa de vidrio especial que había sido convertida en no uniforme mediante un ataque con ácido fluorhídrico. Un haz de luz roja de un láser pulsante de rubí sufría una distorsión al atravesar dicha placa. El haz distorsionado pasaba a un tubo de un metro de longitud, cuatro milímetros de anchura y cuatro milímetros de altura, lleno de gas metano a una presión de 140 atmósferas. La difusión de Brillouin estimulada tenía lugar en dicho tubo; cuando el haz reflejado atravesaba la misma placa de vidrio atacada químicamente emergía sin distorsión. En otras palabras, su estructura era idéntica a la del haz incidente.



2. REVERSIBILIDAD DE LAS ONDAS LUMINOSAS, que tiene consecuencias muy importantes. Un haz direccional ideal (es decir, un haz libre de distorsión y divergencia) sufre una degradación al atravesar una placa de vidrio de espesor no uniforme (*arriba*). El haz se puede regenerar si los rayos individuales que lo integran se invierten y se transmiten hacia atrás a través de la misma placa de vidrio (*esquema inferior*).

La conjugación de fase por difusión



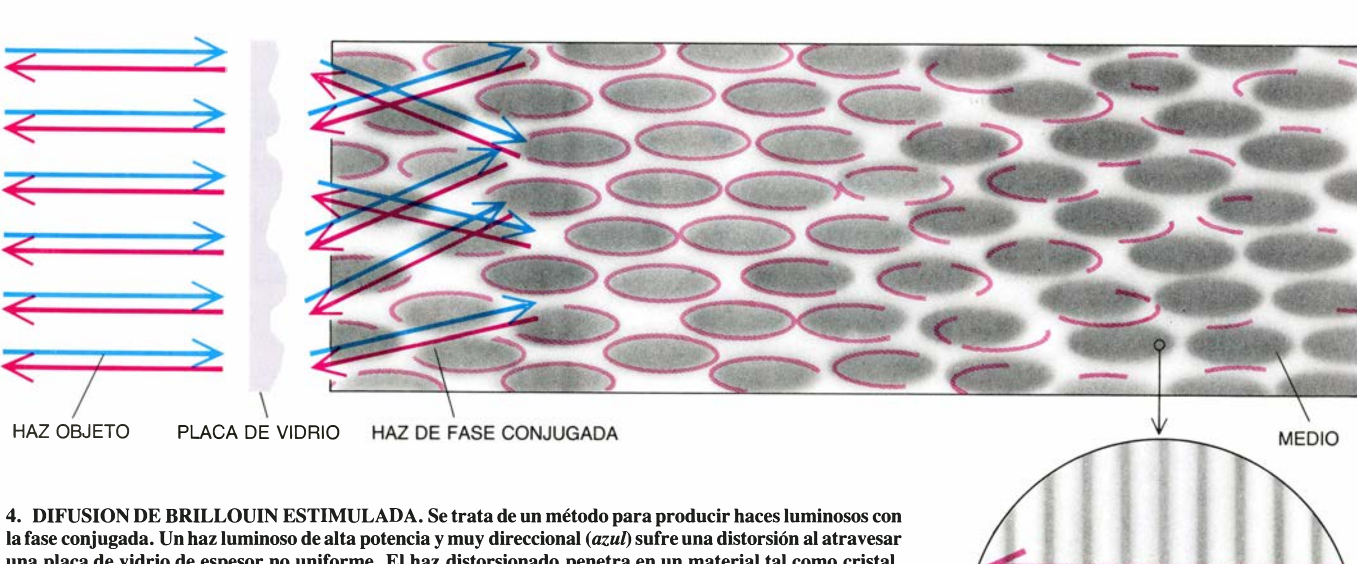
3. ESTA FOTOGRAFIA DE UN HAZ DE LASER sugiere la reversibilidad de las ondas luminosas: si el único dato fuera esta imagen sería imposible afirmar si el haz se mueve de izquierda a derecha o de derecha a izquierda. (La dirección es de izquiera a derecha.) La serie de bandas verticales oscuras aparece como resultado de la interferencia del haz de láser con un haz "de referencia"; estas bandas corresponden a las superficies de los frentes de onda, es decir, a los lugares donde la oscilación es sincrónica. Los frentes de onda de dos ondas mutuamente conjugadas (en nuestro caso, la onda que se mueve hacia la izquierda y la onda que se mueve hacia la derecha) están invertidos en relación con las direcciones de propagación, según se ilustra en el esquema de la derecha).

estimulada se ha llevado a cabo con posterioridad utilizando numerosos medios de difusión y muchos tipos de láser diferentes. La ventaja principal de esta técnica consiste en que sólo requiere una celda llena con un sólido, líquido o gas adecuado. La simplicidad del método de difusión estimulada llevó a decir a Robert W. Hellwarth, de la Universidad del Sur de California, que "a la naturaleza seguramente le gusta el haz de fase conjugada". Es interesante destacar que el propio Hellwarth propuso en 1977 otra forma de conjugar las fases: la mezcla de cuatro ondas.

La mezcla de cuatro ondas es el otro método de conjugación de fase más utilizado en la actualidad. Este método se basa en la interferencia de cuatro haces luminosos en un medio no lineal. Tres de los haces son haces de entrada: uno de ellos es el haz objeto cuya fase se desea conjugar, y los otros dos son haces de referencia. Los haces de referencia, que se mueven en direcciones opuestas entre sí, suelen ser ondas planas y tienen la misma frecuencia que el haz objeto. El haz objeto puede penetrar en el medio considerado en cualquier dirección. El cuarto haz es un haz de salida; es el conjugado de fase del haz objeto y emerge a lo largo de la misma línea que el haz objeto, aunque su sentido de propagación es contrario al de éste.

El haz conjugado se produce como consecuenia de ciertas perturbaciones originadas en el medio material por la interferencia del haz objeto con uno de los haces de referencia. Cuando los campos eléctricos de la ondas oscilan en fase, los campos se suman y la intensidad local de la luz es elevada. Si oscilan con fases opuestas, los campos se restan y la intensidad local de la luz es baja. Las zonas de alta intensidad están intercaladas entre las zonas de baja intensidad. El tamaño, forma y orientación de todas las zonas vienen determinados por las características de los campos que interfieren.

En consecuencia, toda la información sobre la fase del haz objeto está almacenada en lo que se conoce como figura de interferencia, que se manifiesta en el medio en cuestión como



4. DIFUSION DE BRILLOUIN ESTIMULADA. Se trata de un método para producir haces luminosos con la fase conjugada. Un haz luminoso de alta potencia y muy direccional (*azul*) sufre una distorsión al atravesar una placa de vidrio de espesor no uniforme. El haz distorsionado penetra en un material tal como cristal, vidrio, o un líquido o gas comprimido. Este haz se difunde debido a las fluctuaciones de densidad estocásticamente distribuidas (ondas sonoras fluctuantes térmicamente) situadas al final del medio en cuestión, dando lugar a una serie de ondas de configuraciones espaciales variables (*fragmentos rojos curvados*). Cuando la frecuencia y dirección de una onda difundida son exactamente las adecuadas, la onda interferirá con el haz incidente y dará lugar a la formación de más variaciones de presión y densidad en el medio considerado (*barras grises en la región aumentada de la derecha*). Estas variaciones producen la subsiguiente reflexión de una pequeña porción de haz incidente. La porción reflejada interfiere a su vez con el haz incidente, originando nuevas variaciones de presión y densidad. Dichas variaciones de presión y densidad dan lugar, a su vez, a más reflexiones del haz incidente. Las reflexiones aumentan de una manera exponencial, hasta que un haz de fase conjugada muy potente (*en color rojo*) emerge del material. La alta calidad y direccionalidad del haz original se restablece en la transmisión hacia atrás del haz conjugado a través de la placa de vidrio.

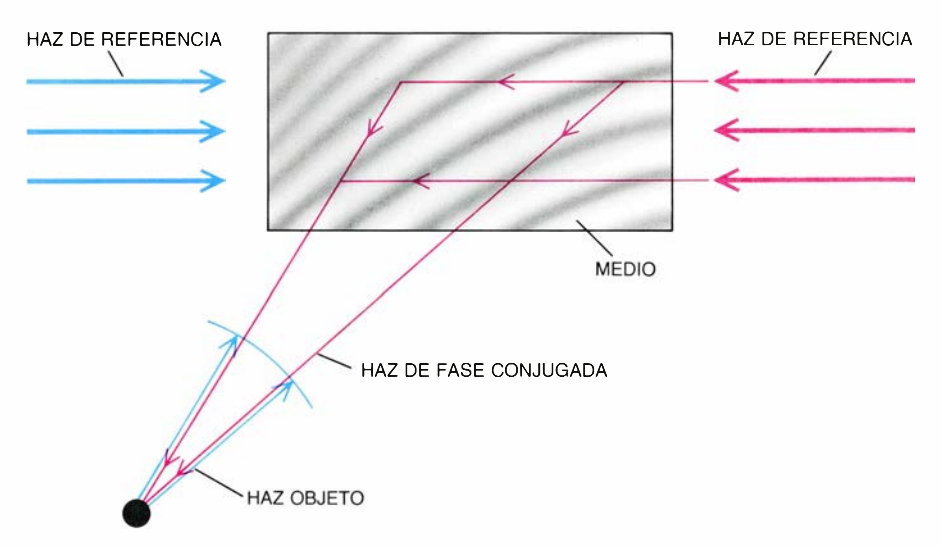
una serie de zonas con índices de refracción diferentes.

El segundo haz de referencia se refleja en las zonas de la figura de interferencia, es decir, el haz lee la información sobre la estructura de la fase almacenada en dicha figura. Dado que la dirección del segundo haz de referencia es opuesta a la del primero, el haz reflejado es el conjugado de fase del haz objeto. Aunque la reflexión en cada zona de la figura es débil, la suma de todas las reflexiones adquiere su importancia y puede transferirse una cantidad considerable de energía desde el haz de referencia hacia el haz conjugado.

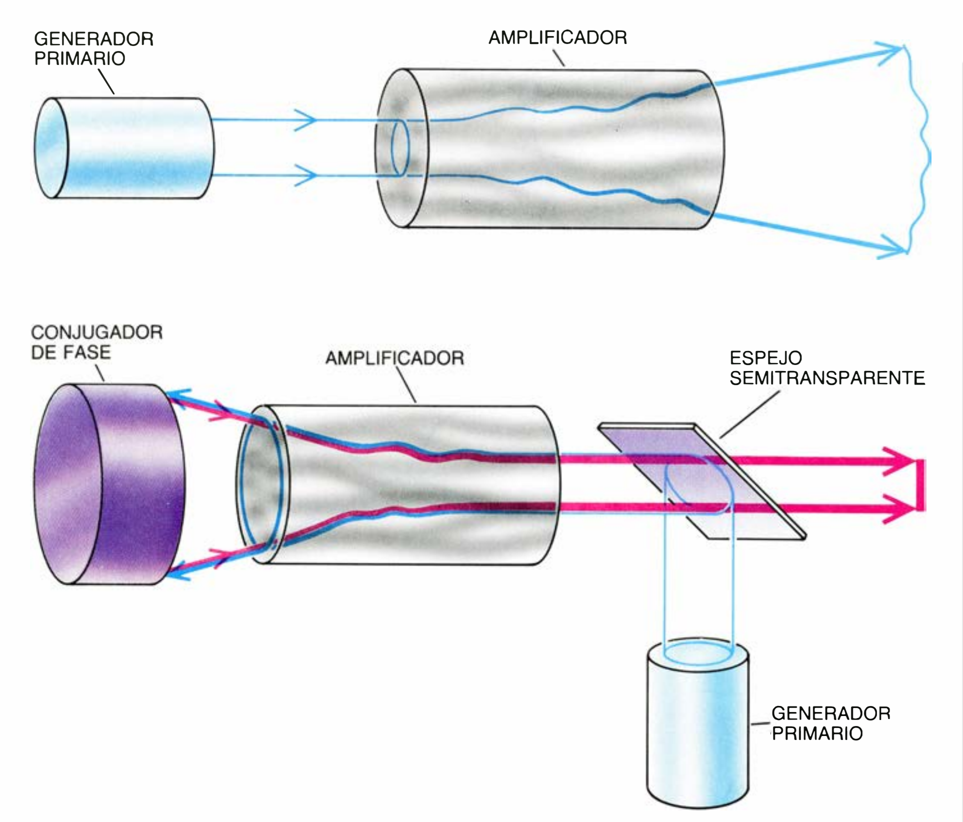
La simetría en la disposición de los haces de referencia sugiere que la figura de interferencia podría crearse también con el segundo haz de referencia y el haz objeto y leerse con el primer haz de referencia. En realidad se forman dos figuras de interferencia en el medio no lineal: cada haz de referencia origina una figura en la que se refleja el otro haz de referencia.

De hecho, lo que acabamos de describir es el registro y la reproducción de un holograma "dinámico". Un holograma es una figura de interferencia formada con haces láser y almacenada en una película fotográfica, lo cual permite producir una imagen tridimensional. La holografía tradicional "estática" requiere tres etapas diferenciadas: en primer lugar, se procede a registrar el holograma iluminando una transparencia fotográfica con la figura de interferencia resultante de dos ondas, un haz objeto y un haz de referencia; a continuación, se revela la película y finalmente se lee el holograma con el mismo u otro haz de referencia. La mezcla de cuatro ondas es un ejemplo de holografía dinámica, debido a que los tres procesos citados –registro, revelado y lectura– tienen lugar simultáneamente. Las variaciones causadas en el índice de refracción desaparecen cuando lo hace la radiación de iluminación, y el holograma está cambiando continuamente en respuesta a las variaciones del haz objeto.

A la vista de lo que antecede, no resulta sorprendente que la idea de la conjugación con cuatro ondas se encontrara ya en los trabajos de los pioneros de la holografía, tales como Dennis Gabor, Yury N. Denisyuk, Emmett N. Leith y Juris Upatnieks. En 1965, Herwig W. Kogelnik, de los Laboratorios de la Bell Telephone, sugirió que las tres etapas de la holografía estática se podrían combinar para dar lugar a un holograma dinámico y, por tanto, producir una conjugación de fase. En 1971 Boris I. Stepanov, Evgeny I. Ivakin y Alexander S. Rubanov, del Instituto de Física de Minsk, junto con J. P. Woerdman, de los Laboratorios de Investigación Philips, en Holanda, llevaron a cabo los primeros ensayos de holografía dinámica utilizando haces de referencia propagándose en direcciones opuestas. Hellwarth inició posteriormente el estudio exhaustivo de la



5. LA MEZCLA DE CUATRO ONDAS constituye otro método para producir haces con la fase conjugada. Este procedimiento se basa en la interferencia de cuatro haces luminosos en un cristal, vidrio, líquido o gas comprimido. Tres de los haces son haces de entrada: uno de ellos es el haz objeto (*representado aquí como una onda esférica*) cuya conjugación de fase se desea, y los otros dos son haces de referencia. El cuarto haz es el haz de salida; su fase es la conjugada del haz objeto. La interacción del haz objeto con uno de los haces de referencia (*azul*) produce una figura de interferencia en el medio material considerado. El segundo haz de referencia (*rojo*) se refleja en la figura de interferencia. Dado que el segundo haz de referencia proviene de una dirección opuesta a la del primero, el haz reflejado es el conjugado de fase del haz objeto. En realidad, todos estos procesos tienen lugar simultáneamente. Más aún, en el medio material se forman dos figuras de interferencia: cada uno de los haces de referencia da lugar a una figura en la que se refleja el otro haz.



6. DIRECCIONALIDAD DE LOS HACES DE LASER: se mejora utilizando la conjugación de fase óptica. La mayoría de los láseres diseñados para producir haces potentes se construyen de acuerdo con el esquema representado en la parte superior de esta figura. Un "oscilador local" o generador primario produce un haz muy direccional a expensas de la potencia de salida. La potencia aumenta cuando el haz pasa a través de un amplificador. Sin embargo, la existencia de irregularidades (inhomogeneidades) en el medio de amplificación produce ciertas distorsiones en el haz. La figura de la parte inferior muestra la forma en que se pueden compensar dichas distorsiones. La luz procedente de un generador primario se refleja con un espejo semitransparente hacia un amplificador. El haz distorsionado que sale del amplificador sufre una conjugación de fase y se devuelve al amplificador. El haz con la fase conjugada se propaga hacia atrás respecto al haz original a través de las mismas irregularidades del medio considerado y, como consecuencia, su movimiento se invierte. El haz de "doble pasada" resultante es de elevada potencia y se encuentra libre de distorsiones.

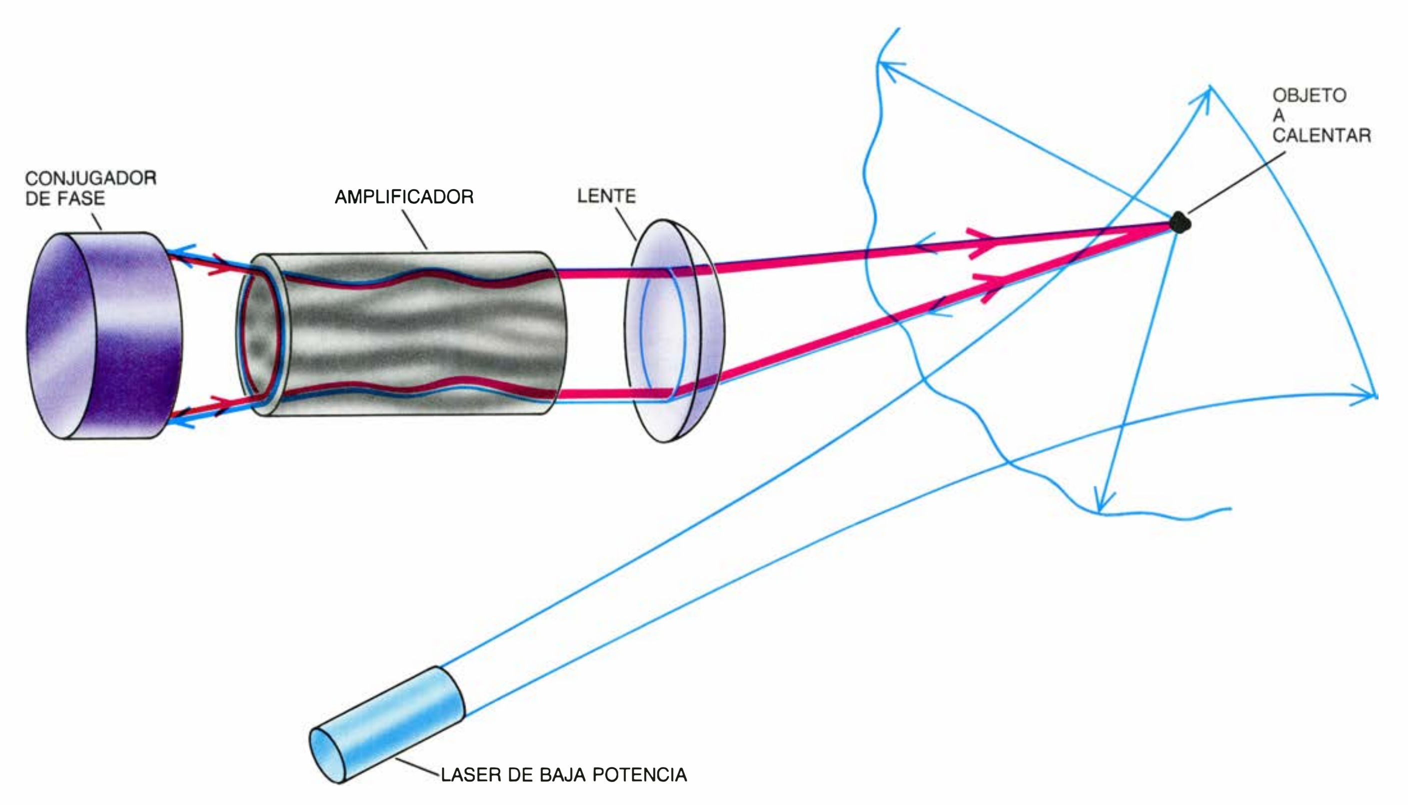
mezcla de cuatro ondas. Este mismo autor describió adecuadamente el fenómeno en términos de óptica no lineal. Entre los investigadores que han desempeñado un papel relevante en el desarrollo de la conjugación con cuatro ondas podemos citar a Ammon Yariv y David M. Pepper, del Instituto de Tecnología de California, y David M. Bloom y Paul F. Liao, de los Laboratorios ATT Bell.

Muchos laboratorios de todo el mundo han utilizado con éxito la conjugación con cuatro ondas. Uno de los atractivos de este método consiste en que, al contrario de lo que sucede en la conjugación de fase por difusión de Brillouin estimulada, no se requiere una potencia mínima de la onda objeto para que se produzca la conjugación. Actualmente se encuentran en estudio otros métodos interesantes de conjugación de fase.

Las posibles aplicaciones de la conjugación de fase óptica son múltiples. La producción de haces de láser altamente direccionales y de radiación autodirigida son dos de las primeras. Esta última aplicación proporciona una forma de calentar un pequeño objeto, cuya superficie puede ser mínima (la millonésima parte de la sección transversal del haz de láser utilizado), sin necesidad de recurrir a un sistema complicado de lentes, espejos y otros elementos ópticos. En el futuro, esta técnica tal vez sirva para calentar un plasma denso e iniciar así un proceso de fusión termonuclear.

Casi todos los dispositivos diseñados para proporcionar haces de láser potentes se construyen de acuerdo con el esquema que detallamos a continuación. En primer lugar, se construye lo que se llama un generador primario ("oscilador local"), capaz de producir un haz altamente direccional a expensas de la potencia de salida. A continuación, el haz pasa a través de un amplificador para obtener la potencia deseada. El amplificador consiste en un sólido o un gas de moléculas muy excitadas. Cuando el haz primario atraviesa el amplificador, estimula las moléculas y da lugar a la liberación de energía en forma de radiación.

Un amplificador homogéneo ideal



7. EN LA RADIACION AUTODIRIGIDA tenemos otra de las aplicaciones de la conjugación de fase óptica. Esta técnica proporciona una forma de calentar un pequeño objeto sin necesidad de recurrir a un sistema complicado de lentes y espejos; podría utilizarse quizá para calentar un plasma denso. Un haz luminoso de baja potencia se dirige hacia las proximidades del objeto a calentar. El objeto difunde la radiación en todas direcciones. Parte de esta radiación atraviesa una lente óptica e incide sobre un amplificador situado cerca del objeto. La radiación aumenta su potencia al atravesar el amplificador. Un conjugador de fase sito al final del amplificador crea un haz "antidistorsionado" y lo refleja a través del amplificador, dirigiendo un haz de láser sobre el objeto.

no distorsionaría la direccionalidad del haz. Sin embargo, la constancia del índice de refracción de un amplificador así tendría que ser mayor que una parte por millón. Difícilmente se puede esperar la obtención de un grado de homogeneidad tan elevado, sobre todo teniendo en cuenta las condiciones de excitación intensa del medio amplificador. El índice de refracción del vidrio, por ejemplo, cambia en una parte por millón cuando la temperatura varía unas tres centésimas de grado centígrado.

Afortunadamente, las distorsiones en la direccionalidad del haz introducidas por un amplificador se pueden corregir mediante un montaje de "doble paso", donde se hace uso de la conjugación de fase óptica. Esta idea fue sugerida y comprobada experimentalmente en 1972 por Oleg Yu, Nosatch y Ragul'skii y sus colegas en el Instituto de Física Lebedev, generando un haz idealmente dirigido con un láser pulsante de rubí. Este haz se hacía pasar por un amplificador de rubí; el haz resultante, potente aunque distorsionado, se sometía a una conjugación de fase óptica. Finalmente, el haz conjugado de fase se devolvía de nuevo al amplificador. Los citados autores encontraron que, durante el segundo paso por el amplificador, el haz consumía casi toda la energía almacenada en las moléculas excitadas del rubí.

Los investigadores han encontrado también algo verdaderamente importante: cuando el haz de fase conjugada atraviesa el amplificador se hace idealmente direccional. La explicación de este fenómeno radica en el hecho de que el haz de fase conjugada se propaga hacia atrás con respecto al haz original, a través de las mismas irregularidades que el láser y, por tanto, "invierte" su movimiento. El haz de fase conjugada compensa no sólo las irregularidades estáticas debidas a los elementos ópticos, sino también las irregularidades dinámicas. La explicación de este fenómeno se basa en el hecho de que el tiempo que la luz invierte en recorrer algunos metros a través del amplificador es del orden de una cienmillonésima de segundo, un valor mucho menor que el tiempo necesario para la excitación y relajación de las irregularidades ópticas en un láser.

La conjugación de fase óptica ha revelado también su utilidad para conseguir radiación autodirigida. Esta idea fue sugerida inicialmente por Kogelnik. Existen varias formas posibles de llevar a cabo esta técnica. Nosotros nos limitaremos a describir sólo un ejemplo. Se empieza por dirigir un haz amplio de un láser de potencia relativamente baja hacia las proximidades del objeto que se desea calentar. El objeto difunde la radiación en todas las direcciones; una parte de esta radiación, después de atravesar algunas lentes ópticas, llega a un amplificador colocado cerca del objeto. A medida que la radiación avanza a través del amplificador, su potencia aumenta. Un conjugador de fase situado al final del amplificador crea un haz "antidistorsionado" y lo refleja sobre dicho amplificador. Como resultado de todo este proceso se obtiene un haz direccional de elevada potencia focalizado sobre el objeto. La focalización del haz sobre el blanco está limitada solamente por la naturaleza ondulatoria de la luz (es decir, por los efectos de difracción) y es independiente de la orientación del sistema focalizador. El blanco parece "atraer" a la radiación amplificada. Nikolay G. Basov y sus colegas del Instituto de Física Lebedev están estudiando la posibilidad de producir la fusión nuclear con láser utilizando la técnica de autodirección de la radiación.

Los investigadores de todo el mundo están explorando actualmente otros muchos aspectos de la conjugación de fase óptica. Al parecer, las aplicaciones posibles de esta técnica son muy numerosas.

Comunicación por vía acústica en el grillo

La capacidad de la hembra para identificar la llamada del macho y localizar la fuente emisora del canto permite estudiar la actividad del sistema nervioso que subyace al comportamiento de ese animal

Franz Huber y John Thorson

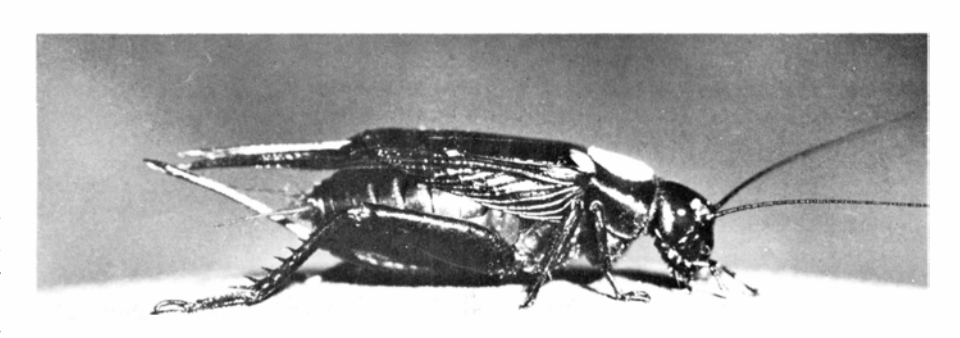
El grillo macho canta frotándose las alas y la hembra le sigue el rastro. Tal interpretación del comportamiento reproductor del grillo parece remontarse a la antigüedad, pero no se verificó experimentalmente hasta 1913, fecha en que Johann Regen, profesor de una escuela superior de Viena, logró que un macho cantara a una hembra por teléfono, invento de reciente instalación. El experimento resultó un éxito: cuando se transmitían los chirridos del macho, la hembra se acercaba al auricular. Pudo así descartarse la participación de estímulos no acústicos, por ejemplo señales químicas, en la respuesta.

Disponiendo de métodos electrónicos y fisiológicos más modernos, nuestras preguntas pueden ser más ambiciosas. ¿Qué características del canto del macho permiten que la hembra localice el foco emisor? ¿Cómo distinguen las neuronas del sistema nervioso central de la hembra ese canto de otros cantos y sonidos? Ante la curiosa disposición de los receptores acústicos del grillo (tiene los oídos debajo las rodillas), ¿cómo determina la hembra la dirección del canto? Más de una docena de laboratorios de todo el mundo tratan, en el presente, de dar respuesta a esas cuestiones. Intentaremos reflejar en este artículo el aroma, el sabor, de algunos de los progresos más recientes. Las respuestas bien merecen el esfuerzo.

Capítulo fundamental de la neurobiología es el intento de explicar el comportamiento de un animal en función de la actividad y las interacciones de algunas de sus neuronas. A menudo, esta investigación se realiza en invertebrados, para sacar partido de su comportamiento estereotipado, de la escasez relativa de células nerviosas y de que muchas de sus neuronas mayores pueden ya identificarse con facilidad en cualquier miembro de una especie, del mismo modo que se identifica fácilmente un diente de una persona cualquiera. Hasta ahora, los éxitos más espectaculares proceden de la porción de salida, o motora, de la organización neuronal. En este sentido, se ha establecido que los complejos movimientos corporales de las sanguijuelas, artrópodos y moluscos están dirigidos por ciertos grupos de neuronas. Los procesos neuronales de los niveles superiores, en la porción de entrada, o sensorial, se han mostrado más escurridizos. Uno de esos procesos contempla una sutil "decisión": el reconocimiento neuronal de la llamada del grillo y el inicio del comportamiento de rastreo de la hembra en la dirección de la fuente del canto.

Es precisamente esta "decisión" lo que pretende explorar la investigación que vamos a exponer aquí. Primero presentaremos un resumen del canto del grillo campestre macho y de lo que sabemos acerca de los órganos auditivos. Seguidamente haremos hincapié en los experimentos neurofisiológicos y de comportamiento que empiezan a mostrar, en detalle, lo que ocurre en el cerebro de la hembra al reconocer el canto. Buena parte de la investigación se ha realizado en el Instituto Max Planck de Fisiología del Comportamiento de Seewiesen, Alemania Federal, en colaboración con nuestros colegas, en particular con Theo Weber, Hans-Ulrich Kleindienst, Klaus Schildberger, David W. Wohlers, Dietmar Otto, George Boyan, Eckehart Eibl, Harald Esch y Leslie Williams.

Los cantos de llamada que estudiamos son los de los grillos que chirrían, a saber, grillos campestres europeos, como el *Gryllus campestris* y el



1. NORIA ESFERICA que permite que un grillo hembra (*arriba*) elija libremente la velocidad y dirección de su paseo, manteniendo a la vez una distancia fija con respecto a los altavoces que emiten los diversos sonidos a ensayar. El grillo camina en lo alto de una esfera de 50 centímetros de diámetro, montada en una cámara anecoica (*derecha*), con un disco de papel de estaño reflectante pegado en el dorso. El papel refleja la luz infrarroja de un foco situado sobre su cabeza, en dirección a unos fotodetectores que registran su posición y envían señales correctoras a los motores que mueven los engranajes de la esfera; dichos motores contrarrotan la esfera para mantener el grillo próximo a la vertical. Las señales correctoras proporcionan registros continuos de la velocidad y dirección del movimiento del grillo. Los altavoces emiten sonidos (cantos de llamada del grillo macho o "cantos" especialmente modificados, sintetizados por ordenador) en diversas direcciones. Al caminar, la hembra expresa su aparente reconocimiento o rechazo de las características del sonido de prueba, rastreando o ignorando los altavoces emisores de tales sonidos. Las fotografías fueron tomadas por Theo Weber, en el laboratorio de uno de los autores (Huber), del Instituto Max Planck.

G. bimaculatus. En los machos de estas especies, las dos alas delanteras poseen una estructura comparable a una lima y un rascador; al cerrarse, las alas frotan entre sí y entran en oscilación a una frecuencia de unos cinco kilohertz (5000 ciclos por segundo). Cada vez que se cierran las alas, por tanto, se emite un pulso de sonido, casi puro, de cinco kilohertz (a menudo denominado sílaba), de una duración de 15 a 20 milisegundos. Oído de cerca, puede resultar dolorosamente estrepitoso. La reapertura de las alas es muda, pero los cierres subsiguientes, y las sílabas generadas en cada uno de ellos, se siguen a intervalos de unos 35 milisegundos, de tal manera que la frecuencia silábica es de unas 30 sílabas por segundo. El tipo más corriente de canto emitido por el macho es una serie de cuatro sílabas, seguida de una pausa breve. Los chirridos se repiten a una frecuencia de dos a cuatro por segundo.

¿Cómo oye los chirridos la hembra? En líneas generales, el aparato auditivo recuerda la organización del humano: un sistema de transformación del sonido, dotado de tímpanos, excita una serie de células receptoras auditivas, que a su vez excitan las neuronas correspondientes. En el grillo, sin embargo, los oídos están debajo de las rodillas de las patas anteriores, a ambos lados del cuerpo: un par de tímpanos en la superficie de cada tibia, cubriendo un conjunto de 55 a 60 células receptoras auditivas.

Los axones, o fibras nerviosas, que parten de las células recorren la pata formando un haz: el nervio auditivo. Su destino es el ganglio protorácico del sistema nervioso central. Los registros de la actividad eléctrica mantenida por el nervio (o de la actividad de las neuronas protorácicas que reciben las señales procedentes del mismo) demuestran que tales células receptoras auditivas reaccionan a sonidos de origen aéreo que van desde unos tres kilohertz hasta lo que serían, para el oído humano, frecuencias ultrasónicas. Evidentemente, la principal población de células receptoras responde mejor a cinco kilohertz, la frecuencia portadora de la llamada del macho.

Hecho destacable es que el oído de



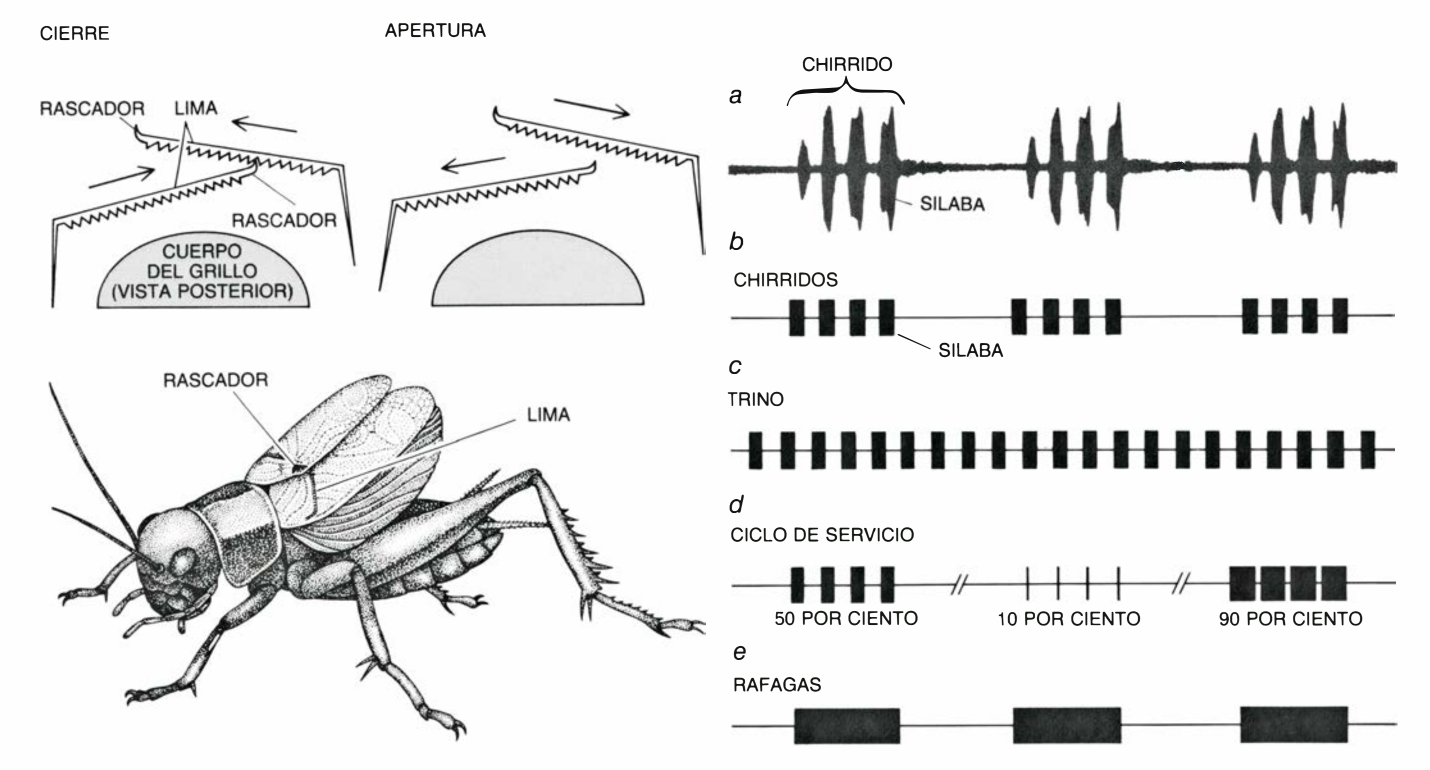
un lado del cuerpo está conectado con el del lado opuesto mediante las ramas inferiores de un tubo traqueal lleno de aire que ya no sirve sólo para la respiración. Las dos ramas superiores terminan en sendos espiráculos: puertas abiertas a ambos lados del cuerpo, al aire que circunda al grillo. En consecuencia, la presión del sonido alcanza los tímpanos por dos vías: la directa, en la superficie externa de los tímpanos, y la indirecta, a través del tubo de aire. Estas relaciones resultan de difícil análisis si empleamos mediciones directas: el pequeño tamaño del tubo traqueal (de 0,1 a 0,3 milímetros de diámetro) obliga a razonar por inferencia. No obstante, muchos estudios, en particular las mediciones por vibrometría de láser realizadas por Axel Michelsen y Ole Naesbye-Larsen, de la Universidad de Odense, Dinamarca, han puesto de relieve las propiedades de tal organización. En estos estudios se ilumina una superficie (en este caso un tímpano de la pata anterior de un grillo) con un haz de luz láser y se calcula la velocidad de la superficie a partir del desplazamiento Doppler que experimenta la frecuencia de la luz reflejada por la superficie.

Los resultados son un tanto sorprendentes. Por ejemplo, el sonido de cinco kilohertz que entra, pongamos por caso, por el espiráculo izquierdo, mueve el tímpano de ese lado (por retropresión desde el interior del tubo traqueal) con mayor energía que un sonido de igual frecuencia e intensidad emitido en la inmediata vecindad externa de dicho tímpano. Al parecer, el tubo traqueal tiene una resonancia próxima a los cinco kilohertz. ¿De qué sirve tan complejo sistema de conductos? Mientras la hembra localiza al macho cantor, sigue una trayectoria en zigzag muy característica, como si obedeciera una sencilla regla: gira hacia el oído que en ese momento reciba el sonido más intenso.

El problema que afronta el grillo es que las patas anteriores, y por tanto los oídos, están separados un centímetro escaso, mientras que la longitud de onda de un tono de cinco kilohertz es de unos siete centímetros. En esas circunstancias, la diferencia de intensidad entre los sonidos que llegan, procedentes de una fuente cualquiera, a la superficie externa de cada uno de los oídos es, a lo sumo, de unos decibelios, lo que no basta para guiar la búsqueda femenina de un macho cantor. Sin duda el sistema de tubos asegura que las diferencias de sonido entre ambos oídos se amplíen por fenómenos de anulación y refuerzo de las ondas en razón de los tiempos de propagación, las resonancias y los desplazamientos de fase que se producen en el seno del tubo. En suma, diríase que el sistema en cuestión, con una resonancia próxima a la frecuencia de la llamada del macho, evolucionó en beneficio de la audición orientada.

La comprensión de la participación de los tubos traqueales en la audición del grillo sigue siendo parcial. Hasta hace poco, por ejemplo, no se sabía si era la deformación timpánica o la presión sónica interna del tubo traqueal (dos fenómenos bien distintos) lo que excitaba los receptores auditivos. (Esas magnitudes pueden discrepar enormemente entre sí en amplitud y fase, según varía la frecuencia y el ángulo de incidencia del sonido.) La incógnita la han despejado Kleindienst y Wohlers, en Seewiesen, en colaboración con Naesbye-Larsen, de Odense. Separaron ambas variables mediante experimentos de anulación sónica.

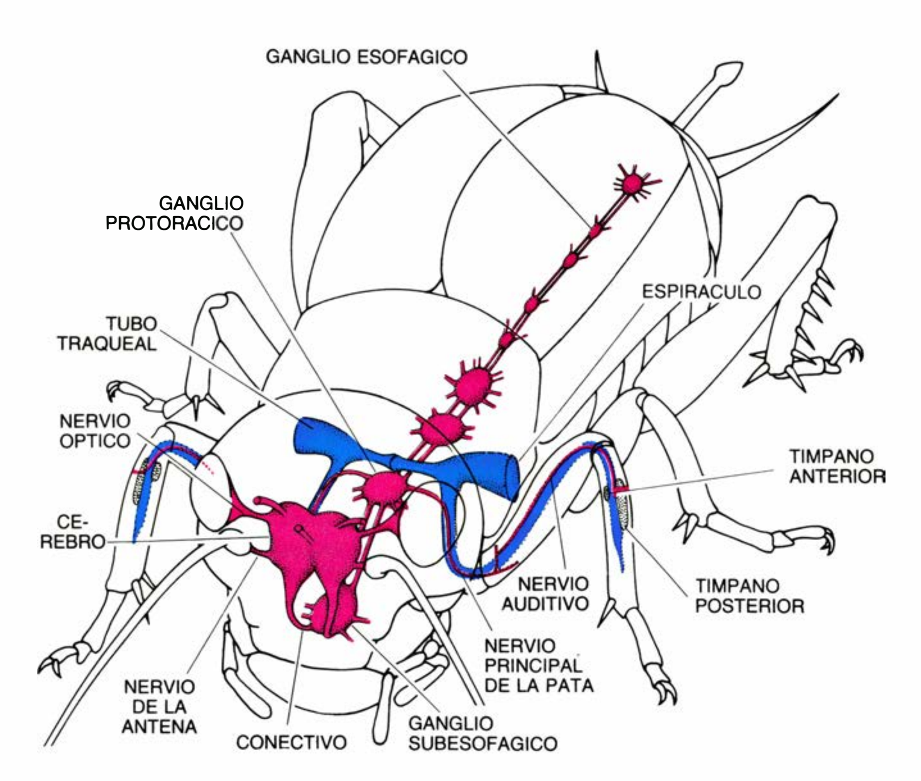
Al efecto se encerraba la pata anterior de un grillo en una pequeña cámara emisora insonorizada (un *patífono*), de modo que el sonido afectara exclusivamente a los tímpanos de dicha pata. A través del cuerpo del grillo, el sonido pasaba por el tubo traqueal y alcanzaba a los tímpanos de la pata opuesta. Simultáneamente se emitía,



2. CANTA EL GRILLO macho frotándose las alas. Los diagramas del lado superior izquierdo muestran el mecanismo productor de sonido: cada ala posee un rascador y una lima; al cerrarse, entran en vibración a una frecuencia de cinco kilohertz (5000 ciclos por segundo). El esquema inferior izquierdo muestra cómo levanta el grillo las alas y las restrega para producir el sonido. El canto (*a*) se compone de "sílabas" de cinco kilohertz. La norma es que haya cuatro sílabas por chirrido, emitidas con una frecuencia de 30 por segundo. Debajo de la pauta temporal del canto verdadero se representan algunos cantos sintetizados por ordenador. El (*b*) reproduce las características fundamentales del canto natural. El (*c*) es un "trino": no divide el canto en chirridos. En el canto (*d*) se conservan los chirridos y la frecuencia silábica, pero se amplía o contrae la duración de cada sílaba, variándose así el "ciclo de servicio", o proporción entre sonido y silencio de cada chirrido. Por último, una pauta de prueba (*e*) integrada por "ráfagas" persistentes de cinco kilohertz de la duración de un chirrido.

desde el exterior, un sonido de amplitud y fase controlables hacia esta misma pata opuesta. Un vibrómetro de láser registraba el movimiento del tímpano posterior, al tiempo que un microelectrodo registraba la excitación de los receptores auditivos midiendo la actividad de una neurona auditiva del ganglio protorácico [*véase la figura 4*].

La idea consistía en ajustar los pulsos externos de cinco kilohertz de tal modo que, al librar pulsos internos y externos sintonizados, el vibrómetro no detectara movimiento timpánico. En esas condiciones, los resultados no dejaron lugar a dudas. Al emitir pulsos, ora externa, ora internamente, el tímpano se movía, los receptores auditivos lanzaban descargas de impulsos nerviosos (claviformes señales eléctricas de la fibra nerviosa) y las neuronas auditivas reaccionaban mediante la producción de varios impulsos. Sin embargo, cuando los pulsos internos y externos coincidían, permaneciendo inmóvil el tímpano, las neuronas auditivas guardaban silencio, aunque la presión sónica interna en la tráquea llegaba incluso a los 80 decibelios. A todas luces, era la deformación timpánica y no la presión sónica local lo que activaba los receptores; los efectos mecánicos directos incidían, es de suponer, en el proceso de transformación del sonido.



3. DISPOSICION DEL TUBO TRAQUEAL (*azul*) que domina el sistema auditivo periférico del grillo y que comunica los dos "oídos", las curiosas estructuras sitas debajo de la rodilla de cada pata anterior. A cada lado del cuerpo, la rama superior, grande, en embudo, del tubo traqueal conduce a una abertura de la superficie del cuerpo: el espiráculo; la inferior desciende por la pata anterior hasta situarse entre los dos tímpanos: membranas auditivas superpuestas a una serie de células receptoras. Gracias a esa estructura, el sonido no sólo llega a cada oído por la vía externa, desde el medio ambiente, sino también por la interna, por la presión sónica del tubo traqueal. El sistema nervioso central del grillo (*rojo*) se compone de una secuencia de ganglios (agregados de neuronas) unidos por pares de conectivos, o haces de fibras emitidas por las neuronas. El ganglio más anterior es el cerebro. Aparece a continuación el ganglio subesofágico, seguido de tres ganglios torácicos y varios abdominales. Las señales auditivas procedentes de las células receptoras sitas debajo de los tímpanos auditivos entran en el ganglio protorácico, el primero de los tres torácicos.

En general, los etólogos estudian el grillo y su capacidad auditiva observando lo que hace la hembra al recibir diversos sonidos emitidos por altavoces. En el experimento clásico, se suelta una hembra en un laberinto en forma de Y, o en una superficie lisa a modo de coso, y se emiten por altavoces diversos sonidos en distintas direcciones. En esos experimentos debe capturarse y reubicarse la hembra después de cada ensayo. Además, cuando el animal ha de escoger entre fuentes de sonido, una salida al azar hacia una de ellas hará que ésta suene más fuerte, simplemente por estar más próxima que antes de moverse el animal.

Se evitan esos problemas si se logra mantener la hembra en el mismo sitio durante las pruebas: en lo alto de una esfera que gire para compensar sus movimientos. El grillo tiene entonces libertad para elegir la velocidad y la dirección del paseo y, aún así, permanece más o menos fijo con respecto a las fuentes de sonido. Ello permite practicar ensayos con el grillo ininterrumpidamente, durante horas; se puede elaborar largas secuencias de sonidos, e idear ensayos para verificar la reproductibilidad de las respuestas y para comprobar si el orden de presentación de los sonidos interviene en los resultados. (Los métodos para estudiar la audición en grillos atados y en vuelo, desarrollados por Andrew Moiseff, Ronald R. Hoy y Gerald S. Pollack, de la Universidad de Cornell, ofrecen algunas de estas ventajas.)

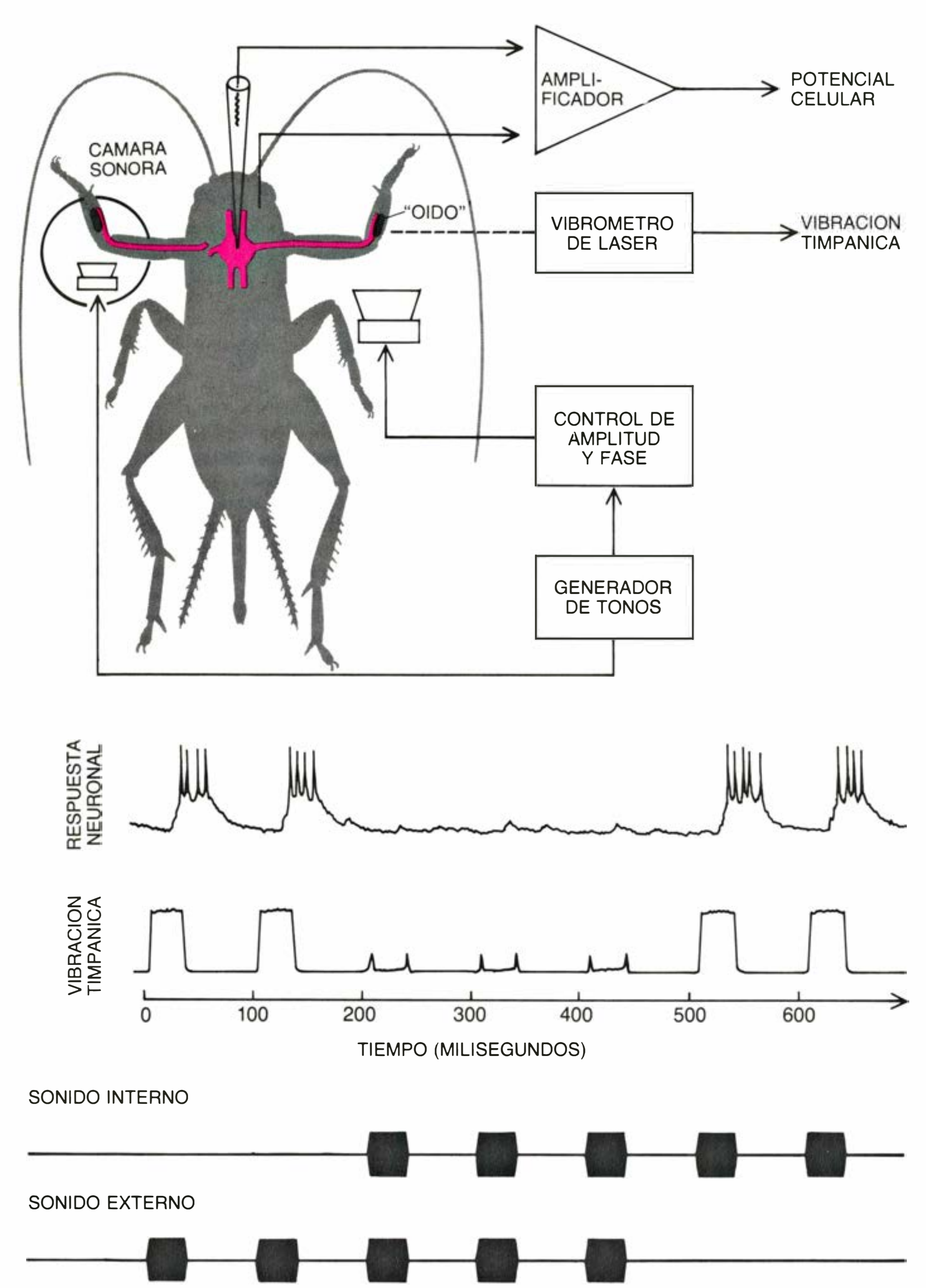
Nuestro dispositivo, que es en esencia una noria esférica, fue diseñado por Ernst Kramer y Peter Heinecke, en Seewiesen, y hoy día se utiliza en varios laboratorios. El grillo se pasea libremente en lo alto de una esfera de plástico de 50 centímetros de diámetro. Se le pega en el dorso un disquito de papel de estaño reflectante y se le hace llegar, desde una fuente situada sobre la cabeza, luz infrarroja de 963 nanómetros, invisible para el grillo. Cuando el animal se desplaza de la vertical de la esfera, unos fotodetectores captan el movimiento por medio de la luz reflejada y los motores de la esfera entran en acción para compensarlo y mantener el animal próximo a la verticalidad.

Para el grillo, ese caminar por la noria esférica en nada se parece a un paseo natural: las aceleraciones correctoras impresas a la esfera alteran inevitablemente las fuerzas de inercia que el animal detecta con los receptores sensoriales de sus patas. No obstante, pueden ajustarse las respuestas de los motores a las órdenes recibidas de modo que no saque más de la mitad del cuerpo del remate mismo de la esfera, sin que por ello sus movimientos parezcan violentos. Siguiendo la trayectoria de las señales de posición y velocidad enviadas por los fotodetectores, se obtiene un registro continuo de la dirección en que el grillo intenta caminar y de la velocidad de su marcha.

El comportamiento del grillo en la noria esférica suele ser bastante diáfano: persistentemente responde de forma afirmativa o negativa a las preguntas acerca de sus inclinaciones. Por ejemplo, cuando suenan chirridos simulados con ordenador por altavoces situados en diversas posiciones en torno a la esfera, la hembra tiende a reaccionar de forma característica. Mediante una serie de paseos cortos, sigue la pista del altavoz que emite el sonido. La precisión aumenta cuando la intensidad del sonido sube de 50 a 70 decibelios. Cuando cambiamos de altavoz, ajusta correspondientemente su dirección. (El experimento se realiza en una cámara anecoica para que el único so-

nido significativo que llegue al grillo sea el procedente de los altavoces.) Por otro lado, la mayoría de las hembras ignora las ráfagas de cinco kilohertz que duran lo que persistiría un chirrido no dividido en la secuencia de sílabas que constituye la pauta natural.

Disponiendo de medios tan formidables para el estudio de las respuestas del grillo hembra al sonido, nos planteamos qué características de la estructura temporal de la llamada del macho podían resultar determinantes a la hora de iniciar el rastreo de la hembra. Entre las posibles candidatas figuraban la frecuencia de los chirridos, la frecuencia de las sílabas y el número de sílabas por chirrido. Investigaciones anteriores ya apuntaban la importancia de la frecuencia de las sílabas. Guiados en parte por esos trabajos, partimos de la hipótesis de que la frecuencia silábica natural (aproximadamente 30 sílabas por segundo) era el único parámetro que necesitaba la hembra para identificar el canto del macho y empezar a buscar el foco emisor.

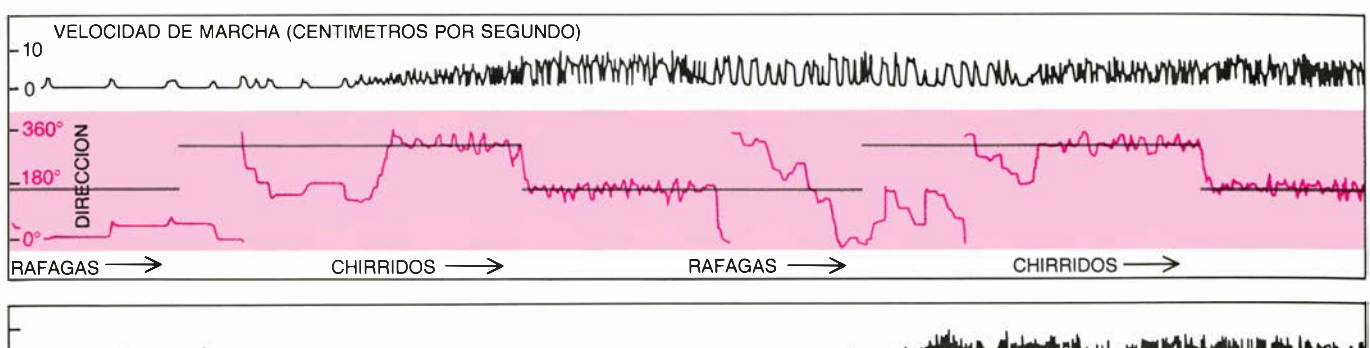


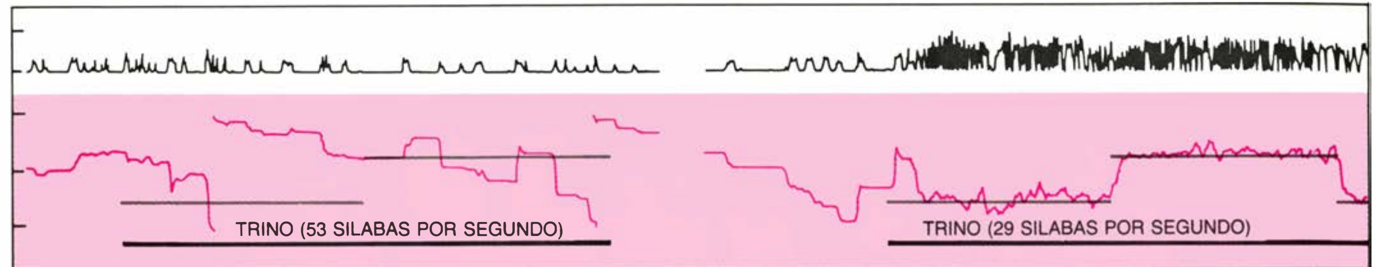
4. EXPERIMENTO DE ANULACION SONORA ideado por Hans-Ulrich Kleindienst, de Seewiesen. Aclaró la función de los tímpanos en la transformación del sonido. Una de las patas anteriores del grillo permanece encerrada en la pequeña cámara representada en el lado izquierdo de la figura. En su interior se dejan oír pulsos de cinco kilohertz; el sonido se propaga por el tubo traqueal y excita los receptores de la pata contraria merced a la presión interna ejercida en los tímpanos. Simultáneamente se emite un sonido externo a la pata derecha de la figura. Un vibrómetro de láser registra el movimiento del tímpano posterior de esta última pata. Se ajustan la amplitud y la fase del sonido externo para que anulen el efecto del interno, con lo cual cesa la vibración timpánica. Se eliminan las señales auditivas del oído de la pata encerrada en la cámara cortando el nervio auditivo. Un microelectrodo intracelular mide la respuesta de una interneurona auditiva del ganglio protorácico. Abajo se muestran los resultados de un experimento. Cuando se envía un sonido, sea interno o externo, al oído de la derecha, el tímpano posterior se mueve y se excita la neurona. Cuando se elimina el movimiento timpánico mediante presentación simultánea de los dos sonidos, la neurona no se excita, aunque la presión sónica en la tráquea interna sea considerable. Muestra así el experimento que la deformación timpánica es decisiva para la transformación del sonido; la duración de la presión sónica endotraqueal no basta para activar las células receptoras del oído. (Dibujos de Patrica J. Wynne.)

En el grillo común europeo, *Gryllus campestris*, la hipótesis sobrevive a un sorprendente número de ensayos. Por ejemplo, la hembra ignoraba chirridos unisilábicos, pero rastreaba los de tres o más sílabas cuando se le presentaban a una frecuencia de 30 por segundo. En otras palabras, la frecuencia silábica es decisiva; el número de sílabas, no.

En un conjunto posterior de experimentos variábamos la proporción entre sonido y silencio en cada chirrido. En un chirrido natural, las sílabas ocupan cerca de la mitad del tiempo, con un "ciclo de servicio" del 50 por ciento. Sintetizando chirridos de cuatro pulsos breves de 5 kilohertz, a una frecuencia de 30 pulsos por segundo, reducíamos el ciclo al 10 por ciento; sintetizando chirridos de cuatro pulsos largos, a una frecuencia de 30 pulsos por segundo, incrementábamos el ciclo hasta el 90 por ciento. La hembra rastreó ambos tipos de canto. Una vez más, la frecuencia silábica se revelaba decisiva; no así la duración de cada una de ellas. Sin embargo, los cantos artificiales con tamañas desviaciones respecto del ciclo de servicio natural exigían un volumen de emisión mayor que el requerido por este último, como sería de esperar, dada la pérdida de 30 hertz de potencia que acompaña a los ciclos anómalos. Por último, sintetizamos "trinos" de sílabas de 5 kilohertz a la frecuencia de 30 sílabas por segundo, sin subdivisión en chirridos. Aunque los informes anteriores indicaban que los trinos no resultan atractivos para los grillos hembra, a no ser que se las hubiera preparado primero mediante la audición de chirridos normales, descubrimos que la mayoría de las hembras rastreaban un trino aun cuando fuera el primer canto del día. La frecuencia de 30 sílabas por segundo parecía ser la señal más importante.

Nuestro empeño en la hipótesis de la identificación por los 30 hertz se situaba en la línea del célebre consejo de Alfred North Whitehead: "buscad la simplicidad, y desconfiad de ella". Y a fe que se presentan complicaciones interesantes. Una minoría de las hembras sometidas a prueba no seguía la pista de los trinos, pero lo hacía de inme-





5. RASTREO DE CANTOS ARTIFICIALES por parte de un grillo hembra en una noria; aparece aquí ilustrado mediante registros típicos de velocidad y dirección. En los dos experimentos, el trazo superior (*negro*) da la velocidad de marcha de la hembra. Cuando rastrea, suele seguir una trayectoria en zigzag, ora deteniéndose, ora moviéndose con velocidades de hasta seis centímetros por segundo. Aparecen dos series de trazos inferiores (*color*): los segmentos horizontales (*gris*) indican la dirección de la fuente sonora, que cambia al conectar un altavoz u otro; las curvas sinuosas (*tono oscuro*) muestran la dirección en que se mueve la hembra. En un experimento (*arriba*), se alternaban secuencias de ráfagas de cinco kilohertz, a intervalos de unos dos minutos, con secuencias de chirridos tetrasilábicos que simulaban el canto natural. La hembra ignoraba las ráfagas, pero rastreaba ávidamente la fuente de los chirridos, cambiando de dirección a la par que lo hacía la del sonido. En otro experimento (*abajo*) se presentaron dos trinos, uno de 53 sílabas por segundo y el otro de 29 sílabas por segundo, que es la frecuencia silábica del canto natural. La hembra ignoraba el trino "erróneo" de 53 sílabas, pero sí rastreaba la fuente del trino de 29.

diato si los interrumpíamos periódicamente. Por otra parte, John Doherty, trabajando con nuestra noria esférica de Seewiesen, ha examinado el comportamiento de los grillos hembra de la especie *G. bimaculatus*, pariente próximo de *G. campestris*. Aunque algunas hembras rastreaban trinos, una minoría seguía ráfagas asilábicas de sonido de cinco kilohertz: refutación manifiesta de nuestra hipótesis de que la frecuencia de 30 sílabas por segundo era el único parámetro determinante en el canto de llamada del macho.

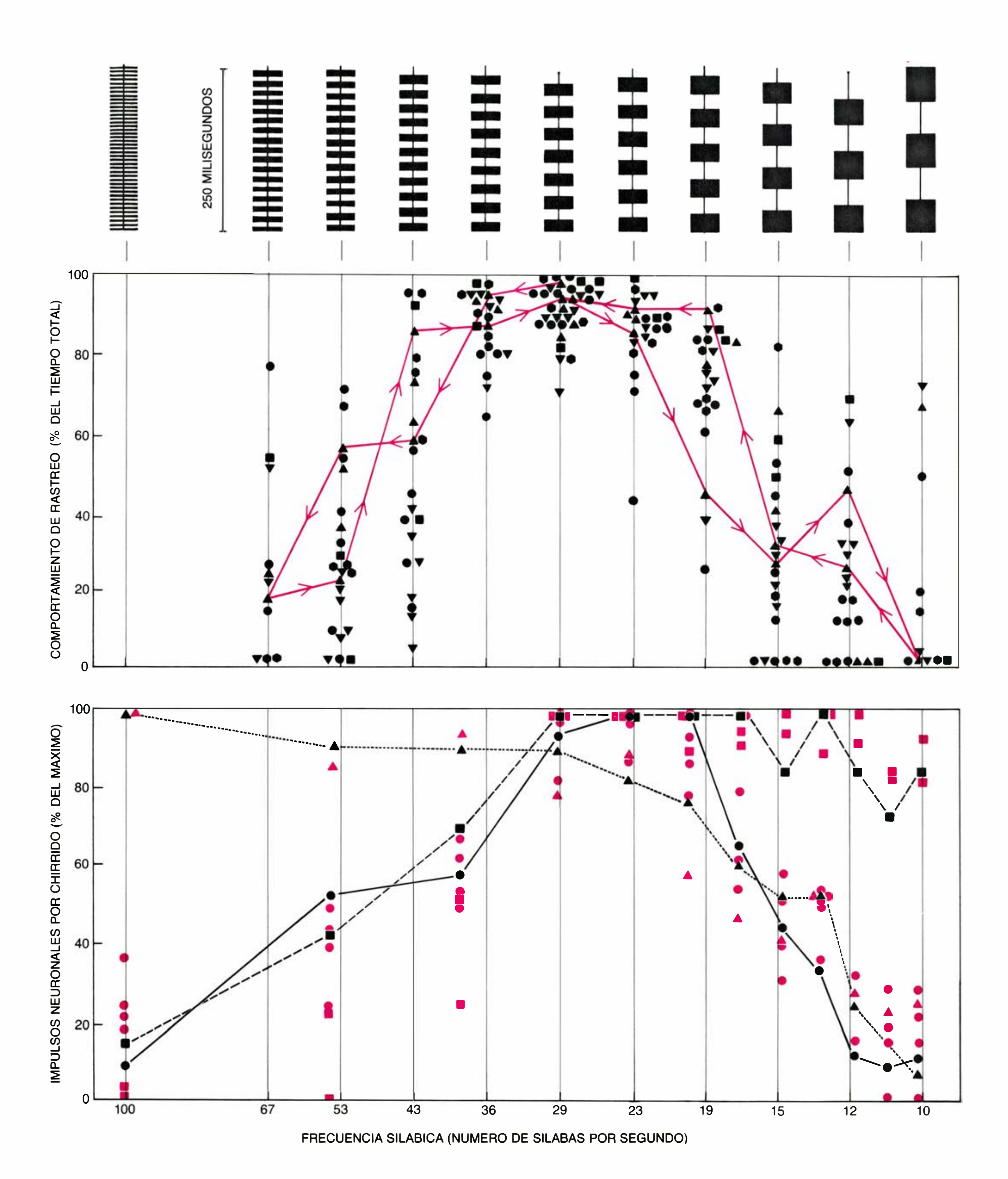
Aún no queda claro si esos resultados constituyen atisbos ocasionales de un complejo proceso de identificación, o son simple reflejo de una carencia anormal de selectividad sensitiva por parte del sistema nervioso central de algunos individuos. Semejantes diferencias arrojan luz sobre un importante punto metodológico: cuando se "llega a conocer" una por una a las distintas hembras durante las largas sesiones de noria en muchas y diversas condiciones experimentales, aparecen diferencias sistemáticas entre los individuos, incluso en su capacidad cualitativa, que no siempre resultan perceptibles en los habituales resúmenes estadísticos de tales experimentos.

En todo caso, es obvio que la modulación del sonido de cinco kilohertz en la frecuencia silábica natural y característica es la forma más segura de llamar a una hembra. En la siguiente etapa de nuestra investigación nos propusimos, de nuevo en colaboración con Weber, capitalizar este hallazgo diseñando un programa de estímulos para localizar las neuronas identificadoras del canto en el sistema nervioso central de la hembra. El problema fundamental de una investigación así es que la mayoría de las neuronas auditivas del sistema nervioso central de la hembra no sólo reaccionan a la llamada del macho, sino también a sonidos que, según sabemos, aquélla no rastrea.

Por tanto, debe ensayarse cada neurona candidata (durante el período, a veces de una frustrante brevedad, en que el microelectrodo está insertado y registra la actividad de la neurona) en un espectro de cantos que incluya los que provocan el rastreo, y otros que no. A tal fin, empleamos una secuencia de cantos sintetizados por ordenador que variaba sistemáticamente la frecuencia silábica, mientras mantenía casi constante la duración, la energía y la frecuencia de los chirridos. El comportamiento de la hembra ante el estímulo de esta secuencia constituye una respuesta acotada: para frecuencias silábicas próximas a 30 por segundo, rastrea casi indefectiblemente, mientras que para frecuencias inferiores a 20 por segundo, o superiores a 40 por segundo, su rendimiento baja de forma drástica.

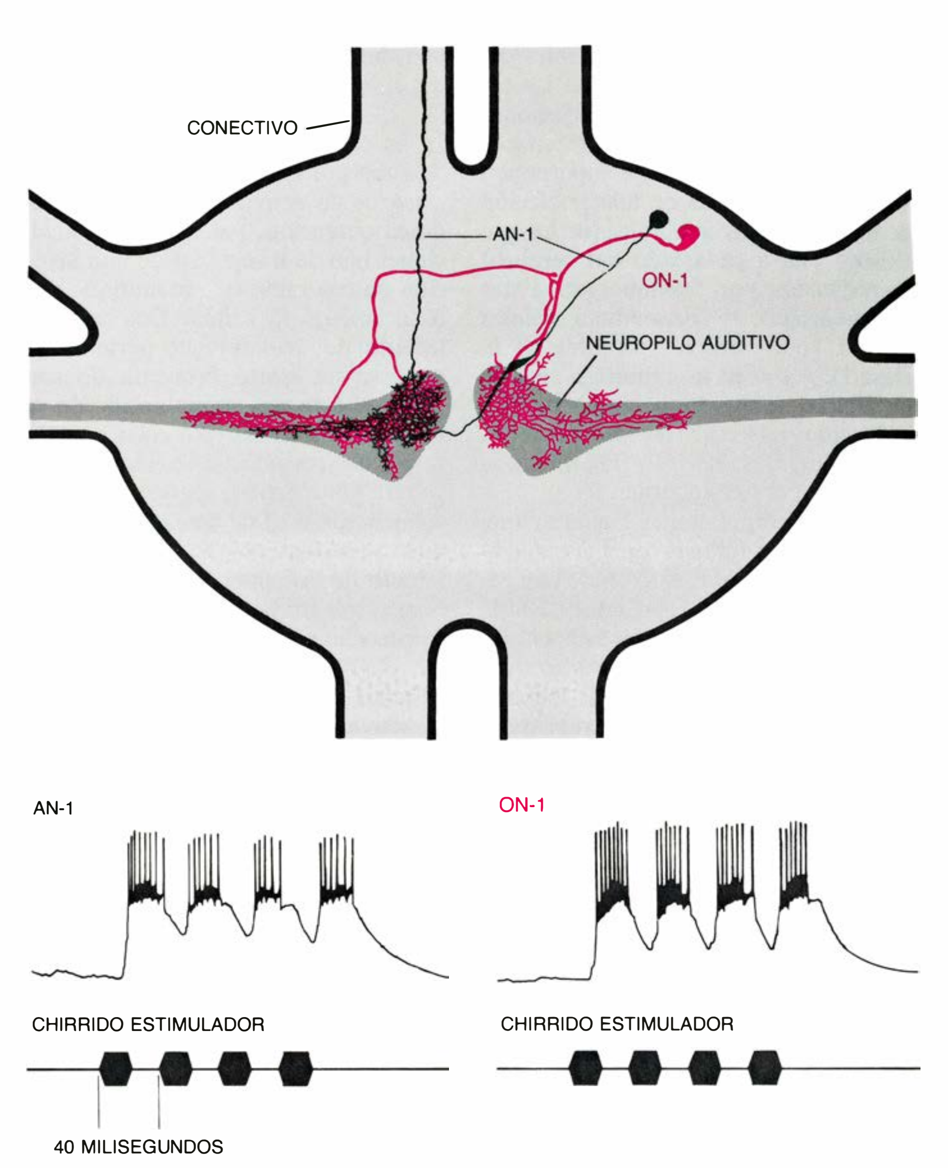
¿Cómo responden las neuronas auditivas centrales a esas secuencias de cantos, correctos e incorrectos? Llegados aquí, se hace imprescindible una somera descripción de la neuroanatomía del grillo. El nervio auditivo discurre por la articulación de la rodilla de la pata anterior, desde la tibia al fémur, y ahí se une con el nervio principal de la pata. Sin embargo, sus axones continúan en un haz compacto, diferenciado, distinguible del resto del nervio, hasta que entran en el ganglio protorácico: complejo neuronal que constituye (entre otras cosas) la primera estación central de tratamiento de la entrada auditiva.

Para determinar el área de distribución exacta de las fibras auditivas se corta el nervio auditivo y se expone el corte a una sal de cobalto, que viaja por las fibras. La precipitación de la sal revela que las fibras se ramifican y terminan en el interior del ganglio, del mismo lado que la pata de la que provienen, sin cruzar la línea media. La restringida región ganglionar donde terminan se conoce como neuropilo auditivo. En él, las ramas terminales de las fibras establecen sinapsis con interneuronas: células nerviosas que procesan la información y la hacen llegar al sistema nervioso central. En los últimos años se han identificado varias de estas neuronas. En realidad, "identificado" es un término técnico de doble significado. Primero, se ha demostrado que las células se caracterizan por una pauta específica de actividad eléctrica y una estructura anatómica también específica. Segundo, se ha puesto de manifiesto que dichas células aparecen sistemáticamente en numerosos individuos de la especie.



6. COMPORTAMIENTO Y ACTIVIDAD de las células cerebrales de los grillos hembra muestran respuestas estrechamente relacionadas a cantos artificiales de diversa frecuencia silábica. Se presentaba a las hembras de la noria esférica una serie de chirridos (*arriba*) que diferían en la frecuencia silábica, pero minimizaban la variabilidad de otros parámetros del canto. La preferencia de las hembras por frecuencias silábicas próximas a 30 sílabas por segundo se plasmaba en su comportamiento de rastreo (*centro*), que muestra una pauta acotada: sensibilidad a un determinado espectro de frecuencias silábicas. Se efectuaron varios experimentos con sendos grillos hembra (*símbolos geométricos*). Los resultados de uno de los ensayos se han unido por flechas de color para mostrar la secuencia en vaivén de las pruebas. En el cerebro del grillo se comprueba que ciertas neuronas específicas (*abajo*) responden a los cantos (en número de impulsos nerviosos desencadenados por cada chirrido) de forma paralela a la respuesta del comportamiento de la hembra. Ciertas células exhiben una respuesta acotada (*círculos*), es decir, responden mejor a frecuencias silábicas próximas a 30 por segundo. Otras tienen una respuesta semiabierta por abajo (*cuadrados*), a saber, reaccionan mejor a, o por debajo de, 30 sílabas por segundo. Aparecen también unas terceras, que muestran respuesta semiabierta por arriba (*triángulos*). Las líneas unen los resultados de una neurona de cada clase. Los datos de comportamiento son de *Gryllus campestris*; las neuronas fueron estudiadas por Klaus Schildberger, de Seewiesen, en una especie próxima, *G. bimaculatus*, que muestra un comportamiento muy similar en todo al de *G. campestris*.

Dos de las más significativas neuronas auditivas identificadas en el ganglio protorácico, y estudiadas ya por diversos investigadores, son las neuronas omega-1, u ON-1, así denominadas por la similitud de su forma con la letra griega omega (Ω). Se presentan como pareja bilateral, la una imagen especular de la otra. Cada célula ON-1 confina sus arborizaciones (la ramificación dendriforme de las protuberancias por las que una neurona despacha o recoge señales) a las dos regiones neuropílicas del ganglio protorácico. Una célula ON-1 sólo recibe señales excitadoras del oído isolateral (que queda del mismo lado) a su cuerpo celular. Es decir, la célula ON-1 cuyo cuerpo celular ocupa la mitad derecha del ganglio protorácico recibe señales excitadoras únicamente del oído de la pata anterior derecha. Y aunque responde en un amplio abanico de frecuencias, la célula ON-1 "copia" punto por punto la llamada del macho, o sea, genera una secuencia de impulsos que imitan cuidadosamente la pauta temporal del canto. En menos de un milisegundo, esas secuencias se transmiten al otro lado del ganglio por el axón, en forma de omega, de la célula ON-1.



7. EN EL GANGLIO PROTORACICO figuran neuronas auditivas que constituyen la primera estación central de tratamiento de las señales que llegan de los receptores del oído del grillo. Las fibras del nervio auditivo (*gris*) reparten señales por una región del ganglio denominada neuropilo auditivo. Una célula nerviosa llamada neurona omega, o célula ON-1 (*color*), recoge dichas señales. Su axón, o fibra nerviosa, en forma de omega, pasa al neuropilo auditivo del lado opuesto del ganglio. A partir de éste, una neurona ascendente, o célula AN-1 (*negro*), envía su axón al cerebro. Los registros de la actividad eléctrica de ambos tipos de célula (*abajo*) demuestran que una y otra copian la pauta temporal de la llamada del grillo macho. Las células ON-1 y AN-1 enantiomorfas (no dibujadas) presentan una estructura similar, que emana de los cuerpos celulares del lado izquierdo del ganglio. Las células se tiñeron por inyección de amarillo lucifer.

Curiosamente, la estimulación del oído del lado contrario al cuerpo celular de cada célula ON-1 inhibe la célula, y esta inhibición la ejerce precisamente su enantiomorfa (como apuntaba el trabajo de Wohlers y demostró recientemente Allen I. Selverston, de la Universidad de California, en colaboración con Kleindienst, en Seewiesen). Así pues, la inhibición es recíproca y, en principio, podría servir para agudizar las diferencias entre las señales auditivas de los dos oídos, incrementando la sensibilidad a la dirección de los sonidos por encima y más allá del refuerzo que proporciona el tubo traqueal. El análisis de otras interacciones neuronales del ganglio protorácico no resulta tan fácil. En primer lugar, un segundo par de neuronas omega, las llamadas células ON-2, se yuxtaponen muy cerca de las ON-1, pero muestran una pauta distinta de activación sensorial. Además, quedan aún otras muchas neuronas pequeñas sin estudiar.

Las líneas de salida que emergen del ganglio protorácico siguen, al parecer, un modelo consecuente con el anterior –decepcionantemente consecuente quizás. En particular, varias neuronas protorácicas grandes, identificadas, presentan axones que viajan hacia el cerebro. Entre esas neuronas "ascendentes", dos tipos destacables, las células AN-1 y las AN-2, se presentan por pares enantiomorfos. Las células AN-1 son muy sensibles a la frecuencia de base del canto de cinco kilohertz, mientras que las AN-2 son sensibles, fundamentalmente, a las frecuencias superiores. A diferencia de las neuronas omega, las AN-1 reciben señales de los oídos del lado contrario; no muestran indicio alguno de recibirlas de los isolaterales. Las arborizaciones dendríticas (recolectoras de señales) de cada célula AN-1 se superponen con las arborizaciones axónicas (transmisoras de señales) de una célula ON-1.

Resulta interesante que entre todas estas células protorácicas identificadas (las neuronas omega y las neuronas ascendentes), no encontremos candidatas a neuronas reconocedoras de la llamada del macho. Todas se limitan a copiar los chirridos –y otros sonidos– y transmiten su pauta temporal al cerebro (o, en ciertos casos, en sentido opuesto: a los ganglios torácicos inferiores). Para estas células, la presentación de un espectro de cantos sintetizados con ordenador, que varíe sistemáticamente la frecuencia silábica, no denota preferencia por las frecuencias próximas a 30 sílabas por segundo: las que mueven a la hembra a rastrear la fuente del canto del macho.

La búsqueda de neuronas reconocedoras del canto nos lleva, pues, al ce-

rebro. En él, las ramas terminales del axón enviado por cada una de las células AN-1 producen arborizaciones densas (descritas por vez primera por Boyan y Williams) que se superponen a las arborizaciones de una población de interneuronas cerebrales (de las que existen una a cada lado del cerebro) identificadas por Schildberger. Estas interneuronas se denominan células BNC-1 (por neurona cerebral de la clase 1). A su vez, las arborizaciones de las BNC-1 se superponen a las de otra población bilateral de interneuronas cerebrales, las BNC-2, identificadas asimismo por Schildberger.

Tales superposiciones sugieren que las señales auditivas que ascienden hasta el cerebro por el axón de una célula AN-1 alcanzan las células BNC-1, y, desde allí, se retransmiten hacia las BNC-2. Nótese a este propósito que, cuando se presenta el canto de llamada a una hembra, y mientras se registra la actividad eléctrica de las células BNC-1 y BNC-2, resulta que la generada por las de clase 1 copia la pauta temporal del chirrido con menor precisión que las neuronas del ganglio protorácico. La copia realizada por las células de clase 2 es menos exacta: los impulsos generados por éstas rara vez muestran una sincronía detectable con las sílabas del chirrido.

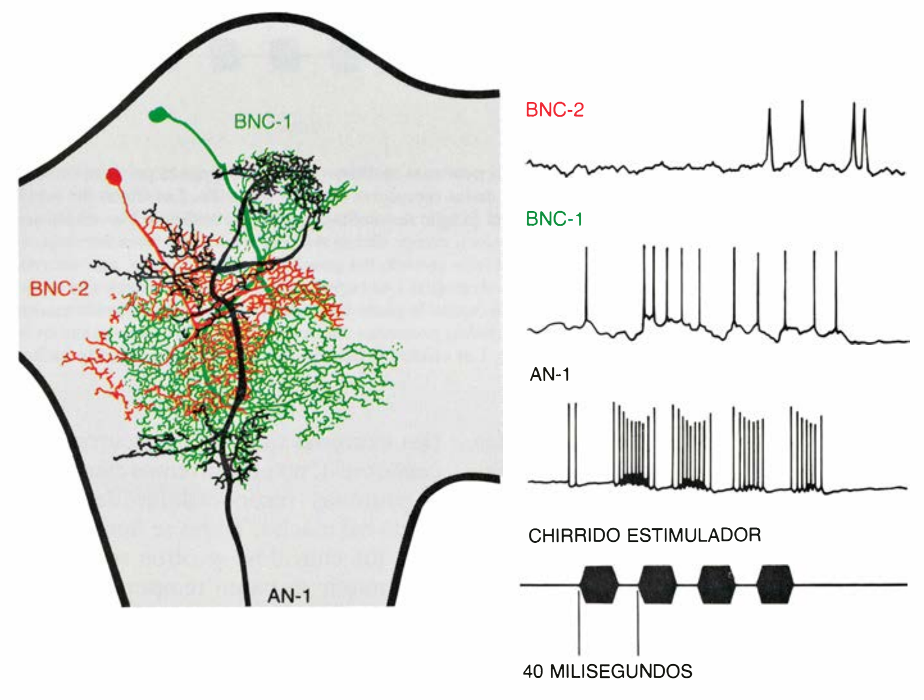
Las células BNC son pequeñas, razón por la cual los registros *in extenso* de su actividad eléctrica son de difícil obtención. Además, tales mediciones han de ir seguidas de una inyección de colorante que identifique anatómicamente la célula. Con todo, el trabajo de Schildberger permite vislumbrar un cierto esquema de conjunto. Existe una subpoblación de células BNC-2 en la que cada neurona responde (en número de impulsos que genera por chirrido, según varía la frecuencia silábica) de una forma acotada que recuerda mucho el comportamiento de la hembra: la neurona responde con preferencia a cantos cuyas frecuencias silábicas se aproximan a las 30 sílabas por segundo, justamente aquellas que, probablemente, habrían de activar el rastreo de la hembra. Que sepamos, el resultado de Schildberger es el primer hallazgo, en un insecto, de una correlación neurocentral clara de un proceso de identificación de pautas temporales.

La población BNC-2 incluye, asimismo, neuronas que reaccionan sólo ante señales con frecuencias silábicas en la banda de reconocimiento, o por encima de la misma. En pocas palabras, están semiabiertas por arriba. Contiene, además, neuronas que responden únicamente a frecuencias silábicas en dicha banda, o por debajo de la misma; es decir, están semiabiertas por abajo. Estos resultados guardan un sorprendente paralelismo con las neuronas acotadas, semiabiertas por arriba y semiabiertas por abajo, descubiertas recientemente entre las neuronas auditivas del cerebro de la rana por Gary Rose y Robert R. Capranica, de la Universidad de Cornell.

No sabemos qué hacen las células acotadas de la población BNC-2 del cerebro de la hembra cuando han identificado la llamada del macho. Las estrechas relaciones entre sus sensibilidades específicas y la banda de frecuencias silábicas que desencadenan el comportamiento de rastreo abundan dentro de la hipótesis según la cual las células BNC-2 activan, de un modo bastante directo, los mecanismos centrales responsables del rastreo. Puede que la actividad de las células BNC-2 conlleve la liberación de un neuropéptido (una breve cadena de aminoácidos) que actúe de mensajero químico, activador del comportamiento de rastreo. ¿O quizás operan las células BNC-2 mediante neuronas descendentes, identificadas también por Schildberger? Sea como fuere, los hallazgos de Schildberger estimulan la especulación en torno a los pormenores del proceso de reconocimiento.

¿Cómo actuaría un identificador neuronal de pautas temporales sencillas? Se han propuesto tres hipótesis distintas. Una de ellas se funda en las sugerencias avanzadas en las dos últimas décadas por diversos investigadores. Entre ellos, Masakazu Konishi, del Instituto de Tecnología de California, que estudió el canto de las aves; Richard D. Alexander, de la Universidad de Michigan, y Hoy y sus colaboradores, de Cornell, que se ocuparon del canto del grillo; y H. Carl Gerhardt, Jr. y Doherty, de la Universidad de Missouri, que hicieron lo propio con el canto de la rana.

En líneas gnerales, afírmase que la hembra dispone de un generador interno de pautas que produce una plantilla con la que comparar las pautas de señales sensoriales arribadas al sistema nervioso central. En el caso del grillo, la hipótesis es muy atractiva. Sabemos que los insectos tienen "generadores de pautas" neuronales para comportamientos rítmicos, como el caminar, el



8. EL CEREBRO del grillo contiene neuronas que parecen intervenir en la identificación (y no mera copia) de la pauta temporal de la llamada del macho. En la figura, el lado izquierdo del cerebro y las dos neuronas de su interior están dibujados según se verían desde la parte superior de la cabeza del grillo. El axón emitido por la célula AN-1 del ganglio protorácico (*negro*) llega de abajo. Sus terminaciones sinápticas delanteras se superponen a las ramificaciones de una población de neuronas, las células BNC-1, de las cuales se ha dibujado una (*verde*). A su vez, las ramificaciones de las células BNC-1 se superponen a las de otro tipo de neurona, las BNC-2 (*rojo*). Los registros de la actividad eléctrica de las neuronas (*derecha*) muestran que no se limitan a copiar la pauta temporal de la llamada del macho. Antes bien, algunas células BNC-1 ofrecen respuestas semiabiertas por abajo para el espectro de cantos sintetizados (*véase la figura 6*); algunas de las neuronas BNC-2 ofrecen respuestas acotadas, y las hay que las dan semiabiertas por arriba. Estos hallazgos indican que las neuronas participan en el reconocimiento de la llamada del macho y activan el rastreo.

vuelo y el canto. Por descontado, el grillo hembra no canta. Sin embargo, comparte con el macho el pertinente aparato motor y neuronal; a veces, cuando se muestra agresiva, mueve las alas de la misma forma que el macho cuando canta. Por otra parte, no está claro cómo se podría comparar una plantilla con el canto de un macho, dada la arbitraria llegada de este último y la ritmicidad de aquélla. Al parecer, se ha de postular un mecanismo desencadenante para la generación de plantillas.

La segunda hipótesis fue propuesta hace más de veinte años por Richard Reiss, de General Precision Inc., de Glendale, California. Recurre ésta a vías neuronales dobles mediante las cuales las señales alcanzan el mecanismo reconocedor. Supóngase que una de las vías retarda cierto tiempo las señales con respecto a la otra. Supóngase, además, que el identificador requiere la coincidencia temporal de las señales procedentes de ambas vías. En esas condiciones, el identificador sólo responderá a ciertas pautas temporales, digamos que aquellas en las que el retardo iguala la periodicidad de la pauta. También en este caso la idea no pasa de mera especulación. Con esta rígida estructura, el identificador aceptaría frecuencias silábicas múltiplos de aquella que está previsto que reconozca. Tales pseudoidentificaciones no se han advertido ni en nuestros datos de comportamiento ni en los registros de las neuronas auditivas del grillo obtenidos por Schildberger.

La tercera hipótesis se apoya en el reciente descubrimiento, en grillos y ranas, de neuronas acotadas en estrecha asociación con neuronas semiabiertas por arriba y por abajo. Quizá, como ha señalado Schildberger, pretenden estas células explicarnos algo acerca del mecanismo de identificación. Es posible que, como principio general, el sistema nervioso identifique las pautas de sonido simples, repetitivas, mediante un proceso en dos etapas. En la primera, dos conjuntos de neuronas determinarían si la frecuencia silábica está por encima de cierto valor (células semiabiertas por arriba) o por debajo de otro (células semiabiertas por abajo). En la segunda, un conjunto de neuronas acotadas, o identificadoras, se activaría sólo si las dos anteriores lo están. En lógica matemática o en análisis de circuitos de ordenador, dicha secuencia corresponde a la operación lógica denominada Y. Los datos anatómicos no contradicen esta hipótesis. Las células BNC-1 parecen enviar señales a las BNC-2 acotadas. Además, las células BNC-1 suelen responder a los estímulos sonoros antes de que lo hagan las BNC-2. Sería interesante saber si existen relaciones equiparables en el cerebro de la rana.

En el estudio que abarca desde el comportamiento del grillo hasta la maquinaria neuronal que lo sustenta, había que determinar cómo las interacciones entre las neuronas pueden producir respuestas semiabiertas a las señales sensoriales. Tropezamos aquí con el desconcierto de la profusión: las interacciones conocidas entre neuronas explican demasiado. Dicho de otra forma, las propiedades neuronales básicas de la excitación y la inhibición ofrecen un amplio abanico de elementos con los que elaborar cálculos. Por ejemplo, el conocidísimo fenómeno de la suma temporal, en el que la llegada rápida de señales excita una neurona, mientras que las más lentas no lo hacen, proporciona un filtro semiabierto por arriba. Digamos sólo que los detalles de las interacciones entre las neuronas identificadas en el cerebro del grillo están evaluándose a la luz de las posibilidades de computación. Todos sabemos cuán difícil resulta comprender en detalle un programa de ordenador, sin documentar, escrito por otra persona.

Mientras tanto, los estudios sobre el comportamiento del grillo en la noria esférica nos deparan continuas sorpresas. Weber y nosotros hemos descubierto que si se sintetizan cantos de llamada a frecuencias de base no naturales la hembra se confunde, rastreando un macho ilusorio en ángulo con la verdadera fuente del sonido. Además, resulta que las hembras que maduran con un solo oído, por un defecto de desarrollo, a menudo pueden rastrear el sonido con bastante precisión. Esos resultados interesan fundamentalmente a nuestra manera de entender la audición orientada del grillo.

Hace doscientos años, el pintor alemán August Johann Roesel von Rosenhof quedó tan fascinado con el comportamiento del insecto que dejó de pintar retratos humanos para recrearse en esos animales. Nos legó un vívido relato del comportamiento del grillo: "la lujuria le tiene [al macho] cantando sin apenas pausa alguna, hasta que finalmente consigue lo que desea; si una hembra anda cerca, marchará en pos del canto". Esa capacidad de la hembra sigue constituyendo un reto para investigadores de la acústica, del comportamiento y de la neurofisiología.

Ciencia y sociedad

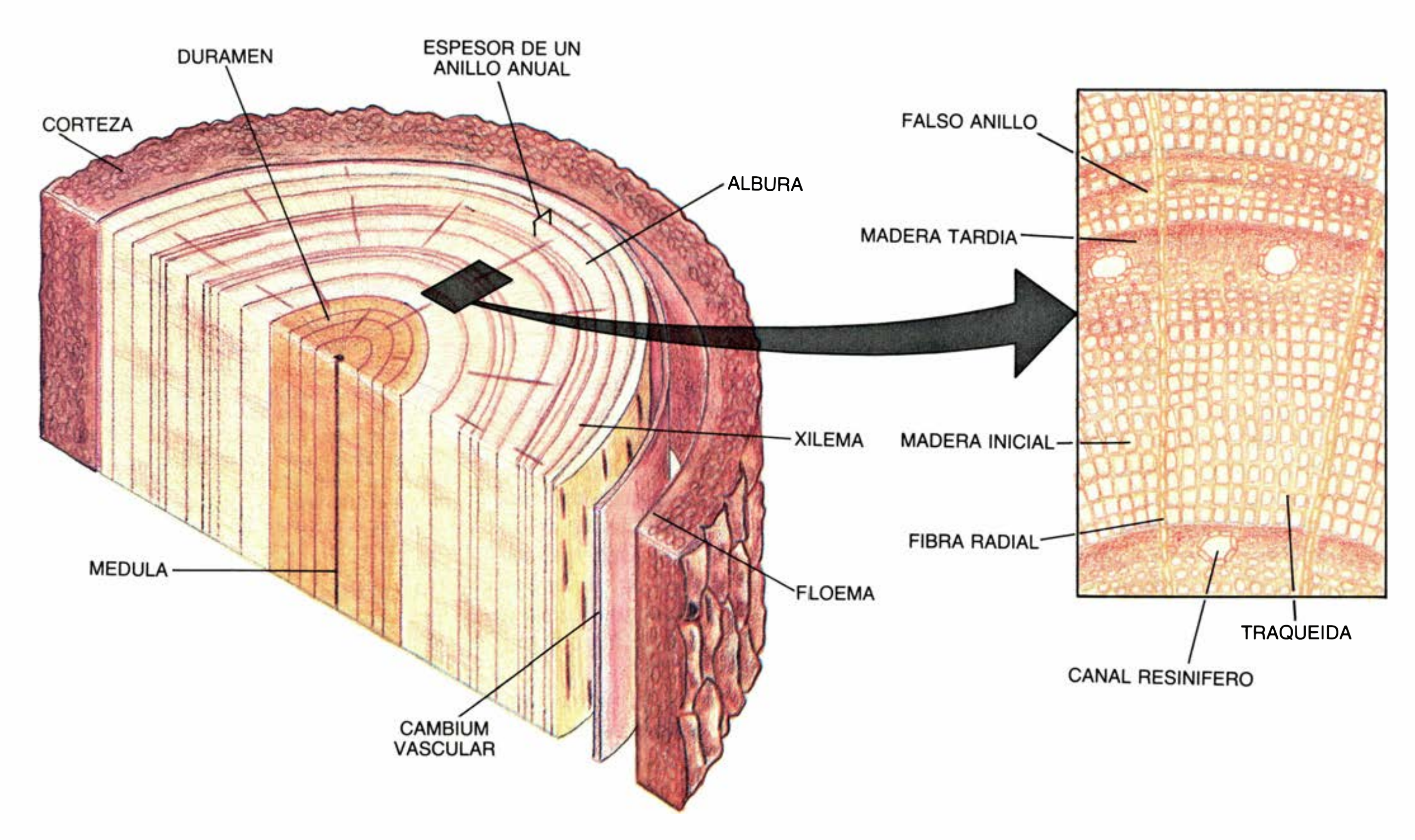
Dendrocronología

El clima es uno de los factores ambientales que tiene mayor influencia sobre los organismos, incluyendo por supuesto al hombre. Toda la actividad de éste se ve limitada, si no amenazada, por el clima: producción agrícola y ganadera, disponibilidad de recursos hídricos, etcétera. Ultimamente, y a raíz de determinados sucesos catastróficos (recordemos la gravísima sequía que afecta desde hace 15 años gran parte de Africa) ha cobrado especial interés el estudio de las oscilaciones climáticas. El fondo de la cuestión estriba en descubrir las posibles periodicidades y adelantarse o prepararse para hacerles frente.

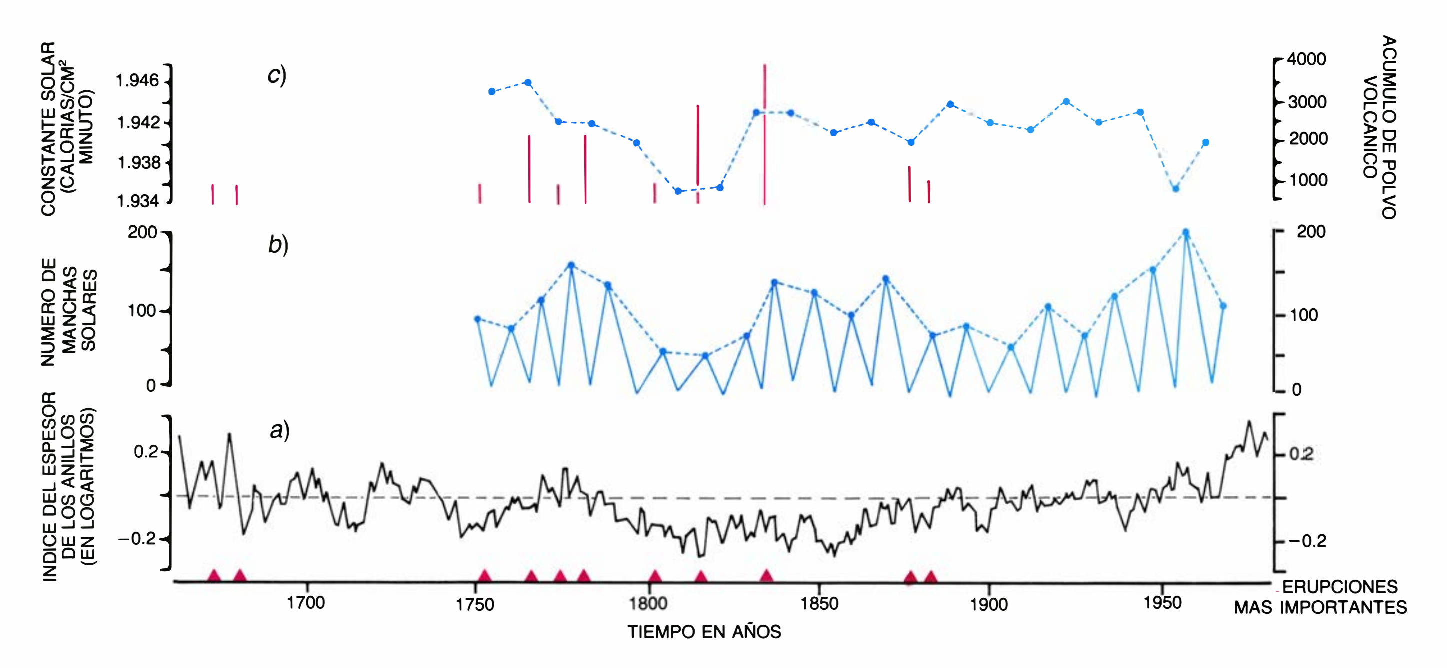
El clima es el resultado o expresión de una serie de factores que operan en escalas de tiempo diferentes. Se distinguen cambios climáticos que presentan periodicidades de millones, miles o cientos de años. A escala humana, son los cambios de corta periodicidad (10-10.000 años) los que revisten un interés inmediato mayor. Entre los factores que modulan estas periodicidades destacan tres: el acúmulo de polvo volcánico, los cambios en la actividad solar y la concentración de anhídrido carbónico en la atmósfera. Ahora bien, mientras los dos primeros llevan actuando miles de años, la importancia del tercero está íntimamente ligada a procesos desarrollados por el hombre durante los últimos 150 años.

El acúmulo de polvo volcánico parece afectar al comportamiento climático de forma directa, ya que las partículas existentes en la atmósfera, tras la erupción, reflejan y absorben buena parte de la radiación solar incidente, con lo que disminuye la proporción de energía que llega a la superficie. Al mismo tiempo, durante la erupción, el azufre pasa a dióxido de azufre, que al reaccionar con el vapor de agua produce una fina capa de aerosol ácido, actuando de pantalla para la radiación incidente. Por otro lado, los efectos de la erupción pueden perdurar varios años. M. R. Rampino y S. Self estiman que una erupción de potencia media, como la de El Chinchón (México) en 1982, puede rebajar la temperatura del hemisferio norte en 0,4°, mientras que otras mayores, como la del Tambora (Indonesia) en 1815, pueden hacerlo en 0,8°. Otro ejemplo lo tenemos en la erupción del Nevado del Ruiz (Colombia) en 1985 con las anomalías térmicas que a buen seguro aparecerán, por su causa, en los próximos años.

El segundo factor consiste en los cambios observados en la actividad solar. Aunque durante mucho tiempo se pensó en la regularidad perfecta del



*1. El crecimiento de los árboles en las zonas templadas forma, durante el período vegetativo, una fina capa celular (vasos o traqueidas) que tiene la misión del transporte de sustancias nutritivas minerales desde las raíces hasta las partes aéreas encargadas de la elaboración de los compuestos orgánicos necesarios para la planta. A principios de este período de crecimiento se forman unas células anchas, con paredes estrechas que se denominan madera inicial, en contraposición con las células estrechas, con paredes fuertemente engrosadas que se originan al final de este período (madera final). Este proceso se repite cada año, y gracias al contraste entre la madera tardía de un año (*n-*1) y la madera inicial del año siguiente (*n*) quedan bien definidos los anillos de crecimiento. El espesor de estos anillos depende de dos grandes grupos de factores: endógenos o propios de cada árbol y exógenos o ambientales. A través de diversos tratamientos matemáticos puede filtrarse la componente interna, con lo que la serie de índices obtenida será un reflejo de la componente externa climática. El espesor de los anillos de crecimiento nos informará, año tras año, de las variaciones que en temperatura, precipitación o radiación, puedan haberse producido durante la vida del árbol.*



2. Comparación entre la serie dendrocronológica obtenida con pino negro, en la Sierra de la Cebollera (Soria), con distintos registros de factores considerados en la actualidad como moduladores del clima: A*) serie de índices del espesor de los anillos de crecimiento (expresado en logaritmos) y datación de las erupciones volcánicas más importantes durante los últimos 250 años (***en abcisas***);* B*) registro del número de manchas solares y envolvente de los máximos;* C*) estimación de la constante solar basada en el modelo de Schneider y Mass. Las líneas verticales (***en rojo***) son una estimación de la potencia de erupción volcánica, calculada como acúmulo de polvo volcánico.*

Sol, conocemos ya la existencia de largos períodos de tiempo con importantes irregularidades. La energía solar llega a los altos niveles de la atmósfera en una cantidad constante (1,94 calorías por centímetro cuadrado y minuto), con ligeras oscilaciones en torno a este valor, variaciones que no han podido confirmarse ya que no sobrepasan el error cometido en la medición de este parámetro. Algunos modelos climáticos admiten que cambios tan sutiles como una disminución del 1 por ciento en el total de radiación solar pueden provocar períodos anormalmente fríos. Las estimas de este parámetro se basan en medidas indirectas efectuadas sobre el número de manchas solares anuales. Esta serie, que se extiende desde 1611, presenta unos ciclos de 11 años, inmersos en otros ciclos más amplios de 22 (ciclo de Hale) en relación con el cambio magnético que presentan.

La concentración de anhídrido carbónico se ha revelado como un factor principal en los cambios climáticos observados en los últimos 150 años. La concentración de este gas se ha elevado, pasando desde 290 partes por millón en 1850 hasta las 320-330 ppm de la actualidad. Este incremento se deriva principalmente de dos actividades humanas: los procesos industriales con utilización de combustibles fósiles y la destrucción de extensas áreas forestadas. Aunque la concentración de este gas en la atmósfera sea pequeña (0,03 por ciento del volumen total) desempeña un importante papel, ya que absorbe la mayor parte de la radiación de onda larga (infrarrojo). Se cree que el aumento en la concentración de CO_2 enrarecerá la opacidad de la atmósfera frente a la radiación infrarroja emitida por la superficie terrestre y que el equilibrio térmico existente en la actualidad se mantendrá a través de un aumento de la temperatura de la superficie y de la baja atmósfera (efecto invernadero).

En 1975, S. M. Schneider y C. Mass propusieron un modelo climático en el que se estimaba la variación de temperatura en superficie desde 1600. Los parámetros tomados en consideración eran la actividad solar y el acúmulo de polvo volcánico. La componente solar se pudo definir a través del número de manchas solares (N), observando que, en el caso de máxima o mínima actividad (N=200) del Sol, los valores estimados para la constante solar eran un 2 por ciento menores que en el caso de que la actividad fuese media (N=80). El acúmulo de polvo atmosférico se obtuvo por combinación de registros históricos y mediciones directas. Las estimaciones concuerdan con episodios climáticos constatados históricamente; por ejemplo, la "pequeña edad glaciar" entre 1650 y 1700, y el período extremadamente frío en los inicios del siglo XIX. Al objeto de verificar la curva de temperatura resultante del modelo, ésta se comparó con registros climáticos directos, como el "índice de rigurosidad invernal de París y Londres" o el de "temperatura anual media para el centro de Inglaterra", e indirectos, como la serie dendrocronológica elaborada para el suroeste de los Estados Unidos o la cronología establecida para los valores de oxígeno-18 en las "muestras" (*cores*) de hielo de Camp Century (Groenlandia). Los datos del modelo registraban, aquí también, notables coincidencias.

De entre todas las técnicas de datación, el análisis de los anillos de crecimiento se ha convertido en una herramienta poderosa. Nos permite registrar variaciones anuales. El espesor de los anillos de crecimiento depende, aparte de determinados condicionantes endógenos controlados hormonalmente, del conjunto de factores ambientales que influyen en el árbol durante su crecimiento. Entre esos factores destacan los climáticos: la anchura de los anillos reflejará, de año en año, las variaciones que puedan haberse producido en temperatura, precipitación o radiación. Esta relación ya había sido observada hace tiempo. Leonardo da Vinci conocía la relación entre anillos anchos y años húmedos; Duhamel, Buffon, Linneo y Burgsdorf determinaron eventos climáticos mediante la observación de series dendrocronológicas. El propio Linneo llegó a definir los anillos de crecimiento como "crónicas invernales" y

Twining los llamaría "registros meteorológicos".

Aunque la relación entre el espesor de los anillos y el clima no es tan sencilla ni directa como pudiera parecer, de lo que no cabe la menor duda es de las posibilidades que su análisis nos brinda. Estudios realizados por nuestro grupo de trabajo con pino negro (*Pinus uncinata* Ram.), en distintos enclaves de la península Ibérica, nos reafirman en esta idea. Amén de caracterizar las variables meteorológicas más importantes en el proceso de formación de madera, de interés para comprender y mejorar la producción de nuestras masas boscosas, se pretendió relacionar el patrón de crecimiento exhibido por la serie dendrocronológica con los factores moduladores del clima.

Observando la serie de anillos de crecimiento establecida resaltaba la aparición de series de anillos estrechos durante períodos térmicamente anormales, por ejemplo, de 1800 a 1820. Por otro lado, algunos años presentaban crecimientos mínimos que, curiosamente, correspondíanse con erupciones volcánicas bien documentadas. Este hecho se había observado antes en series dendrocronológicas elaboradas en el suroeste de Estados Unidos por La Marche y colaboradores. Considerando que una determinada erupción volcánica para un año (n) podría tener consecuencias en el espesor de los anillos de ese año o en años sucesivos (n, $n+1$, $n+2$) se realizó un análisis de contingencia entre ambos grupos de variables. El resultado de este análisis nos confirmó la relación existente entre el acúmulo de polvo volcánico (motivado por la erupción) y una disminución en la producción de madera (registrado como un anillo estrecho); teníamos, pues, una explicación para determinados valores anómalos de la serie dendrocronológica. Sin embargo, el proceso anterior no logra dar cuenta de las tendencias exhibidas a más largo plazo.

Al objeto de caracterizar los factores que pudieran explicar estos patrones de variación, se realizó la comparación de la serie de anillos de crecimiento con el registro de máximos y mínimos de manchas solares recogido desde 1750. El mayor interés se centra en la comparación con los valores de los máximos, ya que en función de su número puede esperarse un aumento o disminución de la radiación que llega a la superficie terrestre. De la comparación visual entre la envolvente de los máximos y los datos dendrocronológicos, resalta la gran similitud observada, quedando bien recogidos períodos anómalos con variación en común; por ejemplo, el registrado en los inicios del siglo XIX. De la misma forma, si los datos de manchas solares se transforman en estimaciones de radiación se advierte una sensible mejora en la comparación, a excepción de los valores recogidos para los últimos 50 años.

La explicación de falta de correlación entre ambas series para este período hay que atribuirla al tercer factor modulador del clima, el acúmulo de CO_2. Se supone que la concentración de CO_2 ha permanecido inalterada durante millones de años y hasta hace poco. Si bien es cierto que no se ha podido comprobar ningún efecto térmico directo durante los últimos años motivado por el incremento observado en la concentración de este gas, no hay que olvidar la importancia que tiene este compuesto sobre la productividad de la vegetación, ya que condiciona en buena parte la actividad fotosintética.

Aunque existan pequeñas diferencias regionales y temporales, la proporción de CO_2 permanece constante en un rango vertical durante los primeros 80 kilómetros. Sin embargo, a medida que subimos en altura la densidad del aire decrece por efecto de la presión, con lo que la presión parcial de dicho gas también disminuirá por unidad de volumen. Lo que se puede esperar, por tanto, es que el incremento observado durante los últimos 150 años tenga efectos muy diferentes sobre la vegetación, según la altitud a la que ésta se encuentre. La importancia relativa del incremento en la concentración de CO_2 será mayor en la respuesta fisiológica de especies que crecen en grandes altitudes, como es el caso del pino negro. Este factor parece ser la causa del aumento de producción observado durante el último centenar de años en coníferas que medran en diferentes regiones montañosas.

Pese al carácter provisional de estos datos, hemos de insistir en el interés de esta línea de investigación que permite, a través del análisis de la información contenida en los anillos de crecimiento, adentrarnos en el conocimiento del clima del pasado. (Ricardo Génova.)

Nobel de química 1985

Hace dos décadas, descifrar la estructura de una sencilla molécula bioquímica a partir de datos de cristalografía por rayos X exigía años de trabajo; hoy se logra en pocos días. Subyacen a ese avance las técnicas matemáticas denominadas métodos directos. Citando "la gran importancia de esa metodología para la investigación química", la Real Academia Sueca de Ciencias otorgó el premio Nobel de química de 1985 a los desarrolladores de los métodos directos, Herbert A. Hauptman, de la Fundación Médica de Buffalo, Estados Unidos, y Jerome Karle, del Laboratorio de Investigación Naval, también norteamericano.

Los haces de rayos X que atraviesan el cristal de una substancia recogen datos de su estructura. Los planos de los átomos o moléculas del cristal dispersan las ondas, que interfieren entre sí y generan un patrón de dispersión que refleja la geometría a gran escala del cristal. La información sobre la estructura de las unidades del cristal está contenida en las intensidades relativas y las fases de las ondas dispersadas en cada punto del modelo.

No existen obstáculos para la medición directa de la intensidad de la radiación, pero cuando Hauptman y Karle desarrollaron sus métodos, en la década de 1950, se consideraba inaccesible la información sobre la fase. Para resolver la estructura de una molécula, por simple que fuera, los investigadores solían "decorar" la molécula con átomos de un elemento pesado, por ejemplo, oro; los átomos pesados dispersan los rayos X con más intensidad que los átomos ligeros de las moléculas biológicas. Al comparar los datos obtenidos con varias muestras, embutidas en distintos elementos pesados, cabía inferir las fases de las ondas dispersadas por los demás átomos.

Apoyándose en propuestas anteriores de que las intensidades relativas de las manchas del modelo de difracción contienen información sobre la fase, Hauptman y Karle desarrollaron un procedimiento matemático para extraer esa información. Proposición fundamental de sus métodos directos es que si tres manchas intensas del patrón de difracción ocupan posiciones cuyas coordenadas se anulan por suma, sus fases relativas también se cancelan mutuamente. Los cálculos efectuados a partir de muchos tríos de manchas arrojan las fases probables de un número significativo de ondas. Análisis matemáticos posteriores llevan a una solución probable de la estructura de la molécula completa. Sujetos a controversia en un principio, los métodos directos constituyen hoy un procedimiento habitual en la determinación de la estructura de pequeñas moléculas, así hormonas, vitaminas y fármacos.

No son adecuados, sin embargo, esos métodos para el análisis de proteínas u otras macromoléculas, en cuyo estudio se mantiene aún la aplicación de la técnica de los elementos pesados. Los

avances que se registran en las fuentes de rayos X, en la obtención de datos y en la capacidad de cómputo, probablemente permitan en un futuro próximo la aplicación de los métodos ideados por Hauptman y Karle a moléculas biológicas complejas. "Son grandes las expectativas", afirma William L. Duax, de la Fundación Médica de Buffalo, "de que en esas técnicas se funde una revolución en el estudio de las estructuras proteicas".

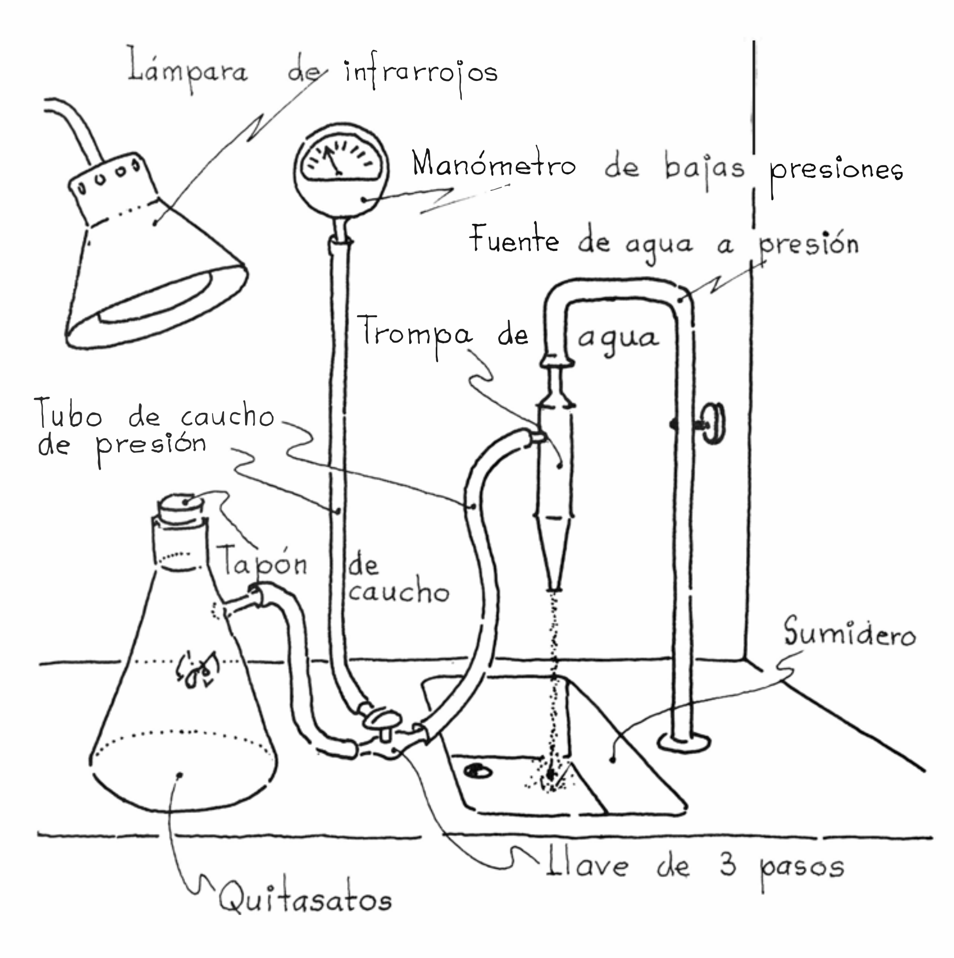
El vuelo del abejorro

Sikorsky, famoso diseñador aeronáutico, había mandado colocar el siguiente rótulo en el vestíbulo de su oficina técnica: "el abejorro, según los cálculos de nuestros ingenieros, no puede volar en absoluto, pero el abejorro no lo sabe y vuela". Son bastantes los estudios acerca del vuelo de muchos insectos y todos tropiezan con enormes dificultades cuando tratan de explicar los mecanismos de sustentación a través de la dinámica de fluidos estacionarios. Veamos algunos ejemplos.

Torkel Weis-Fogh escribía hace once años en *Scientific American* que el aspecto más importante (la sustentación) de esos insectos y otros voladores dependía en buena parte de efectos aerodinámicos no estacionarios, hasta el presente desconocidos, que para ellos son beneficiosos y no un estorbo como lo serían en los aeroplanos fabricados por el hombre. En otro trabajo sobre *Haplothrips verbasci*, Arnold M. Kuetbe se expresaba en términos parecidos: "El desconocimento de los detalles acerca del mecanismo de vuelo, a tan bajo número de Reynolds, indica la necesidad de extensas observaciones durante el vuelo para determinar el movimiento de las alas-barra y de los cilios y, asimismo, la necesidad de profundizar en el estudio de los detalles por medio del microscopio electrónico, y también de mediciones encaminadas a precisar las propiedades físicas del conjunto de cilios..."

Podríamos añadir numerosísimos testimonios más. El lector encontrará un claro planteamiento del problema en el artículo que Joel G. Kingsolver publicó en estas mismas páginas sobre la ingeniería de la mariposa (octubre de 1985). Entre otras cosas, describía las dificultades halladas en el complejo vuelo de estos insectos, muchas de ellas insalvables, recurriéndose de nuevo a las hipótesis de Torkel Weis-Fogh.

He venido investigando, desde hace años, empírica y teóricamente, en un nuevo planteamiento de la dinámica, de la que la clásica sería un capítulo restringido. Entre otras cosas nos abre la posibilidad de que exista sustentación y propulsión incluso en ausencia de atmósfera. En efecto, ¿cómo explicar, desde el punto de vista dinámico, el vuelo de los insectos? Evidentemente, no es razonable en el marco de la dinámica newtoniana en el que la conservación del momento lineal, en un sistema aislado, prohíbe este tipo de sustentación y propulsión.



Vuelo del insecto en el vacío. Montaje del experimento.

En el campo de la cosmología se detectaron, ya hace muchos años, las insuficiencias teóricas de la mecánica newtoniana en sus axiomas de partida. Así, el "primer principio" afirma que un punto material (o un sistema) aislado describe una trayectoria recta y uniforme; pero el movimiento debe estar referido a unos ejes coordenados inerciales, que son externos al punto material (o sistema) considerado, con lo que el aislamiento que se postula queda en entredicho, pues nos conduce a la afirmación contradictoria de que un sistema aislado goza de la propiedad de no estar aislado. Este es "el punto más débil del soberbio edificio de la mecánica newtoniana" (P. Hoenen, 1948). Es necesario corregir este primer principio afirmando que no existen sistemas inercialmente aislados.

Con este nuevo punto de partida, unido al axioma de la conservación de la energía, se inicia la elaboración de esta nueva dinámica comenzando por el caso más sencillo en el que la energía potencial es conservativa, para generalizarlo, en un segundo paso, al caso no conservativo. Nos conduce al sorprendente resultado de que, además de las fuerzas de inercia newtonianas, en las que sólo intervienen las aceleraciones de las partículas y sus respectivas masas, existen otras fuerzas de inercia –hasta ahora desconocidas– en las que intervienen, además, las velocidades de las partículas, cuya masa puede comportarse como no constante en el caso no conservativo. Estas fuerzas resultan ser isomórficas con la "fuerza de Lorentz" del electromagnetismo, cuyo origen es puramente empírico.

En el caso conservativo, la partícula queda afectada de una fuerza adicional: fuerza de arrastre, que se superpone a la newtoniana y es normal a la trayectoria; goza de la cualidad de cambiar de signo cuando el punto material invierte el sentido de su movimiento sobre la trayectoria, lo que implica que ésta sea irreversible en general. Tenemos un ejemplo en el cometa Halley, que podría presentar asimetría a su

paso por el perihelio, es decir, el arco de entrada puede no ser idéntico al de salida.

Vayamos a la observación empírica. Nos servirá de banco de prueba el abejorro, *Bombus terrestris*. El utillaje con el que contemplé la sustentación "anormal" del insecto en el vacío constaba de una bomba de vacío, un recipiente de cristal, una válvula y un manómetro (*véase la ilustración adjunta*). La bomba de vacío puede ser de las conocidas con el nombre de "trompas de agua", empleadas para filtrado en los laboratorios de química. No debe emplearse ningún otro tipo de bomba por una razón muy simple: es del todo necesario mantener la presión parcial de vapor de agua a la temperatura ambiente, con el fin de que el insecto no se hinche ni se deforme, como ocurriría si usáramos otro tipo de bomba aunque el vacío que se consiga sea más elevado. Además, gracias a la rapidez y eficacia, el insecto mantiene sus posibilidades de acción en el vacío durante uno o dos minutos como máximo. A la temperatura ambiente de 15 grados Celsius, se consigue un vacío de 10 tor (13 milibares) que frente al valor normal de la presión atmosférica (1013 milibares) supone un vacío del 98,7 por ciento.

Como recipiente de cristal transparente donde colocar el insecto es muy adecuado un "quitasatos" de 1000 centímetros cúbicos, con un cerramiento hermético de caucho y una salida lateral donde enchufar el tubo de presión, también de caucho, para hacer el vacío en el momento oportuno. No interesan recipientes mayores que el indicado, con el fin de minimizar el tiempo de vaciado –unos diez segundos– y así disponer del máximo período de observación. El insecto se introduce por la abertura superior y luego se cierra herméticamente.

La válvula, de las llamadas de "tres pasos" –las hay muy simples y baratas fabricadas en cristal–, intercalada en la tubería de presión, conecta la bomba de vacío con la salida lateral del "quitasatos". Esta válvula nos permite restablecer la presión atmosférica en el recipiente, después de haber hecho el vacío, sin necesidad de desconectar la bomba y mantener el vacío por tiempo indefinido una vez realizado. Sirve también para comprobar el nivel de vacío logrado mediante un manómetro conectado en derivación. Por lo que al manómetro para bajas presiones se refiere, son muy seguros los de mercurio o bien los manómetros de precisión con lectura digital.

Es sabido que los insectos activan su capacidad de vuelo si alcanzan la temperatura adecuada. (Bueno será, pues, colocar, cerca del recipiente, una lámpara tipo "flexo" que además de iluminar proporciona el suficiente calor, por radiación.)

Los resultados observacionales a que se llega son sorprendentes: durante uno o dos minutos el insecto sigue volando, o arranca a volar, sin diferencia perceptible con el vuelo a la presión atmosférica normal, incluso el tipo de vuelo en flotación, sin movimiento en sentido vertical ni horizontal. La posición de las patas del insecto es la habitual en vuelo, esto es, recogidas y plegadas hacia atrás.

La frecuencia de aleteo es una característica de cada insecto que varía entre límites muy estrechos: alrededor de 300 hertz para el abejorro y 150 hertz para la mosca. La sustentación varía linealmente con la densidad del fluido, de modo que un vuelo en estas condiciones –si lo quisiéramos explicar aerodinámicamente– supondría que el insecto es capaz de levantar un peso casi 100 veces superior al propio; lo que no parece científicamente admisible.

En el caso del vuelo de los insectos, el problema es, en general, no conservativo y de forma análoga aparecen fuerzas, hasta el presente no conocidas, responsables de su sustentación y propulsión –en ausencia de atmósfera– que permiten la explicación del hecho empírico que presentamos. Esto es consecuencia de que en este nuevo planteamiento dinámico no rigen, en general, las leyes de conservación del momento lineal y del momento angular.

La dinámica clásica sigue siendo perfectamente aplicable a aquellos casos en que el sistema se comporta como si estuviera inercialmente aislado, por simetrías, aceleración tangencial nula, órbita circular, etc., o bien las nuevas fuerzas resultan despreciables respecto a las debidas exclusivamente a las masas y aceleraciones de las partículas.

La irreversibilidad termodinámica, "el extraño y molesto segundo principio" (J. Merleau-Ponty) incompatible con la dinámica clásica (teorema de Misra-Poincaré), queda de manifiesto como corolario del nuevo planteamiento dinámico, así como el dualismo partícula-onda. Las ecuaciones de Maxwell del electromagnetismo son deducidas como un caso particular límite. Es de notar que D. W. Sciama en 1953, y Félix Tisserand 80 años antes, y más recientemente Brans y Dicke, intentaron un proceso inverso: construir una teoría de gravitación isomórfica con el electromagnetismo de Maxwell. (Juan Rius-Camps.)

SIDA y sistema inmunitario

El virus del SIDA altera el crecimiento y la función de los linfocitos T4, *glóbulos blancos fundamentales para el sistema inmunitario. El dominio de su modo de acción redundará en beneficio de nuevas terapéuticas*

Jeffrey Laurence

En 1981, el Centro para el Control de las Enfermedades de Atlanta reconoció los primeros casos de una nueva enfermedad fatal, que dio en llamarse síndrome de inmunodeficiencia adquirida (SIDA). Sus víctimas morían de diversas infecciones y procesos malignos, entre otros, una neumonía producida por *Pneumocystis carinii* y el sarcoma de Kaposi, cáncer de las células que tapizan el interior de los vasos sanguíneos. Sufrían también infecciones de "oportunistas", microorganismos ubicuos que de ordinario no son patogénicos. De hecho, las infecciones y cánceres padecidos por los afectos de SIDA sólo se habían observado con anterioridad en personas nacidas con ciertos defectos de su sistema inmunitario y en pacientes cuya inmunidad había alterado la quimioterapia contra el cáncer o las drogas inmunosupresoras administradas en los trasplantes de órganos. Al parecer, el SIDA mataba a sus víctimas destruyéndoles el sistema inmunitario.

A partir de su identificación, la incidencia de la enfermedad se ha elevado vertiginosamente; sólo en Estados Unidos se registraban cerca de 14.000 casos a finales de 1985. Lo que se sabe acerca del SIDA ha aumentado en pareja proporción. Françoise Barré-Sinoussi, Jean-Claude Chermann y Luc Montagnier, del Instituto Pasteur de París, y un grupo encabezado por Robert C. Gallo, del Instituto Nacional del Cáncer de los Estados Unidos, identificaron el agente causal de la enfermedad, un virus de la familia de los retrovirus. Lo lograron, independientemente, en 1983 y 1984, respectivamente. El grupo francés lo denominó LAV (siglas, en inglés, de virus asociado a una linfadenopatía) y los investigadores americanos HTLV-III (por virus *T*-linfotrópico humano de tipo III).

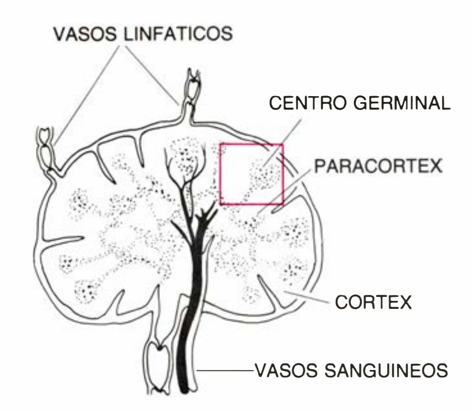
La definición de los grupos de más alto riesgo ante la infección ha mejorado progresivamente: varones homosexuales y bisexuales, drogadictos que se inyectan, compañeros sexuales de individuos incluidos en los grupos de alto riesgo e hijos de madres de alto riesgo. Los receptores de transfusiones de sangre y de productos sanguíneos han contraído también el mal, pero el cuidadoso rastreo de la sangre donada, para detectar en ella el agente infeccioso, ha reducido drásticamente ese riesgo. La enfermedad no ha mostrado signos de extenderse más allá de esos grupos, excepto a dianas predecibles, como mujeres inseminadas artificialmente con esperma de donantes infectados, lo cual indica que el virus sólo se transmite, de ordinario, a través de la sangre, o por contacto sexual. Todas las pruebas epidemiológicas señalan que ni el agua, alimentos, insectos, o contacto casual propagan el SIDA.

Se está adquiriendo también un conocimiento más preciso de los mecanismos por los cuales la enfermedad inutiliza el sistema inmunitario. Esos avances quizá permitan aminorar los efectos de la enfermedad, llegando incluso a prevenirla y curarla. A su vez, plasman la complejidad del sistema inmunitario. Según parece, el colapso total de las defensas inmunitarias de las víctimas del SIDA arranca en gran parte de un único defecto: la reducción del número de los linfocitos *T*4 (uno de los muchos tipos celulares que constituyen el sistema inmunitario) y un cambio en su función.

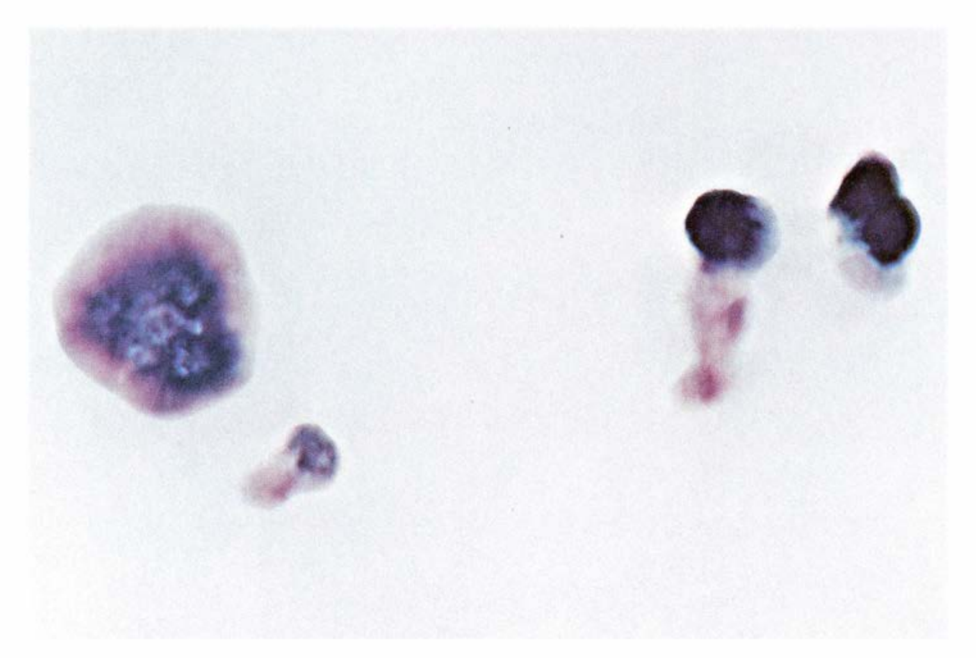
El sistema inmunitario

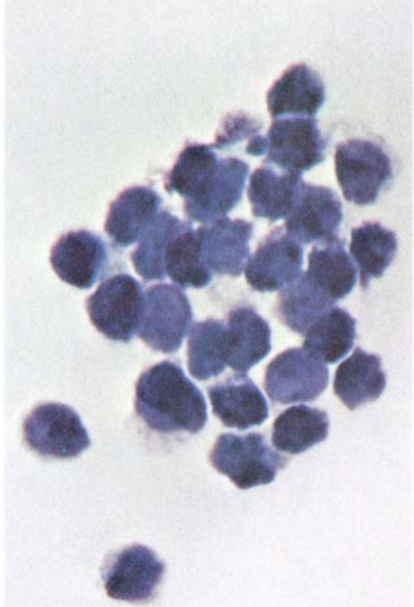
¿Cómo logra el virus del SIDA, dañando sólo uno de sus hilos, alterar la trama entera del sistema inmunitario? La respuesta se halla en la compleja red de interacciones que se dan entre las diferentes clases de células sanguíneas involucradas en la inmunidad. El sistema inmunitario es un mecanismo de defensa flexible, pero sumamente específico, que mata microorganismos y las células que estos infectan, destruye células malignas y retira sus restos. Distingue estas amenazas de los tejidos normales gracias a su capacidad para reconocer antígenos, moléculas foráneas, y monta una respuesta que se amolda a la naturaleza del antígeno.

A lo largo de toda la vida, el cuerpo se enfrenta a miles de antígenos diferentes. Las células del sistema inmunitario pueden reconocer y responder prácticamente a todos ellos; están di-



1. UN DESBARATAMIENTO del nódulo linfático caracteriza al SIDA y al conjunto de síntomas conocido por complejo asociado al SIDA (ARC), uno de sus precursores. En las microfotografías de fluorescencia de cortes de nódulos linfáticos se muestran las estructuras enmarcadas en el esquema (*arriba*): un centro germinal y parte del córtex y paracórtex circundantes. En un nódulo normal (*arriba*) los glóbulos blancos de la sangre denominados linfocitos *T*8 (*naranja*) pueblan el paracórtex, pero no alcanzan el centro germinal, donde forman un entramado regular los linfocitos *B* y las células estructurales del nódulo, las células reticulares dendríticas (*verde*). El corte de un nódulo hinchado procedente de un paciente de ARC (*abajo*) contiene muchas más células *T*8; aquí invaden el centro germinal, cuya estructura regular aparece destruida. No se sabe cómo el virus del SIDA produce esos cambios en la arquitectura del nódulo, pero quizá estén relacionados con la deficiencia inmunitaria; parece que las membranas de la superficie de las células reticulares dendríticas atrapan y retienen proteínas foráneas que desencadenan una respuesta inmunitaria. El desbaratamiento del nódulo linfático bien pudiera servir de clave precoz de la infección por el virus del SIDA. George Janossy, de la Royal Free Clinic de Hampstead, Inglaterra, es el autor de las imágenes.

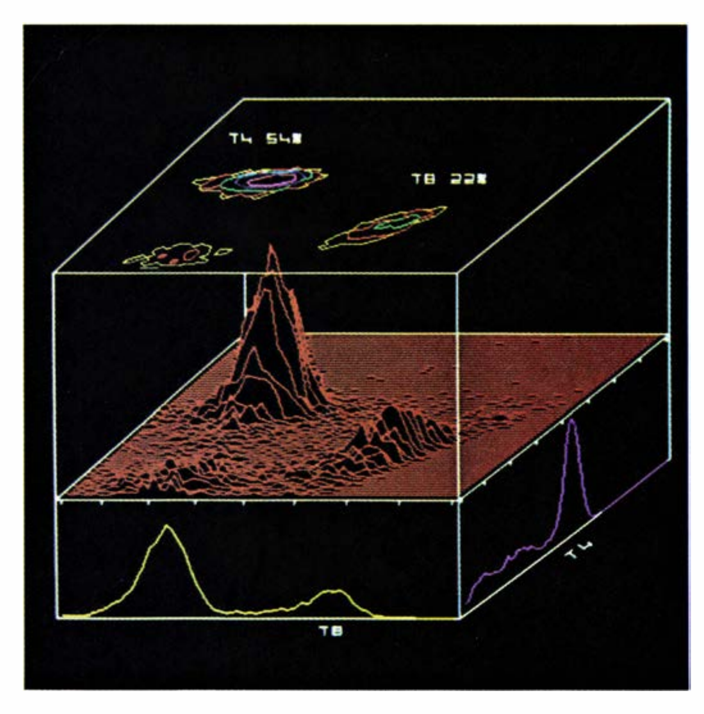


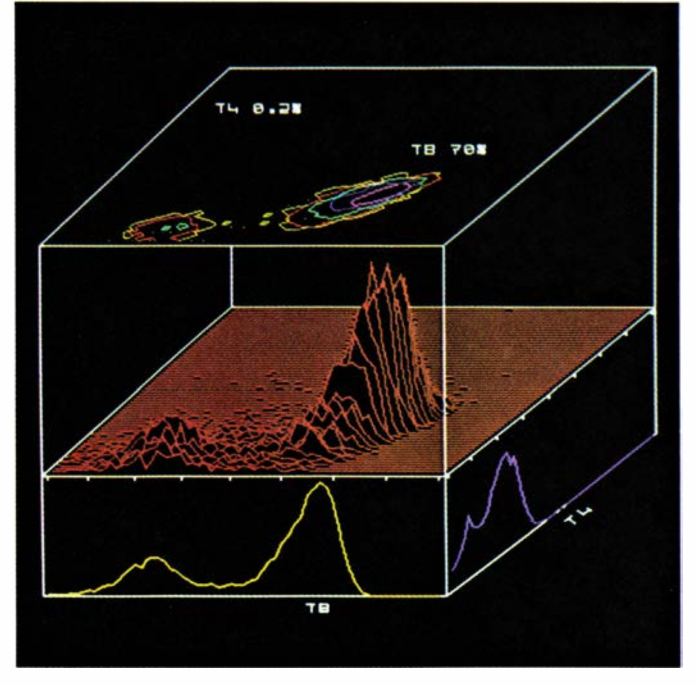


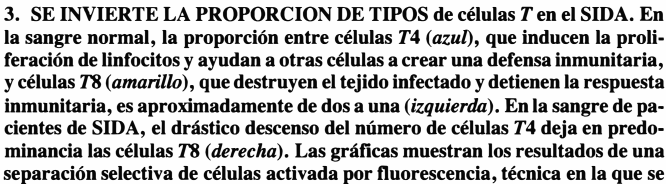
2. CELULAS *T* INFECTADAS (*izquierda*). Se distinguen bien de las células *T* normales, pues el virus del SIDA insta la fusión de sus membranas, formándose complejos multinucleados denominados sincitios. El virus infecta primordialmente células *T*4 que, si bien ofrecen el mismo aspecto que otras células *T*, se caracterizan por su función y por marcadores bioquímicos. La fusión de células infectadas con células normales podría propagar el virus a otros tejidos.

vididas en millones de clones, cada uno de ellos compuesto por una o más células especializadas en reconocer a uno o varios antígenos diferentes. Además de esa diferenciación por su especificidad al antígeno, las células se separan en varios grupos a medida que van madurando a partir de células precursoras comunes. Los tipos de células pueden distinguirse de acuerdo con los contextos en los que reconocen al antígeno, por las características moleculares de sus membranas de superficie y por el papel que desempeñan en la respuesta inmunitaria.

En el más amplio sentido, los glóbulos blancos de mayor importancia en la respuesta inmunitaria, los linfocitos, se encuadran en dos clases: las células *B*, que se desarrollan en la médula ósea, y las células *T*, que también proceden de la médula ósea si bien completan su desarrollo en la glándula del timo. Las células *B* son la fuente de los







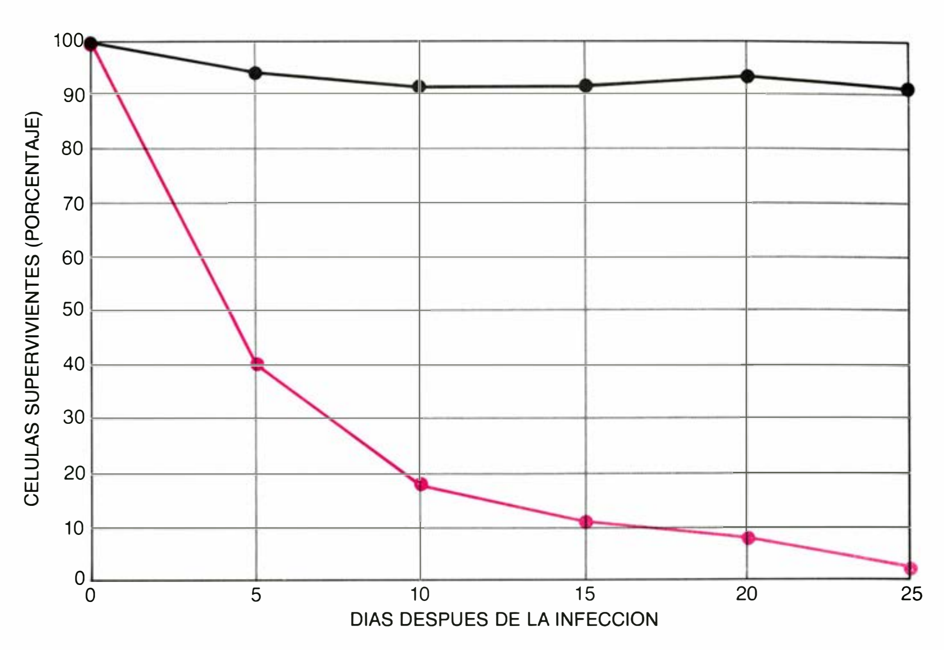
3. SE INVIERTE LA PROPORCION DE TIPOS de células *T* en el SIDA. En la sangre normal, la proporción entre células *T*4 (*azul*), que inducen la proliferación de linfocitos y ayudan a otras células a crear una defensa inmunitaria, y células *T*8 (*amarillo*), que destruyen el tejido infectado y detienen la respuesta inmunitaria, es aproximadamente de dos a una (*izquierda*). En la sangre de pacientes de SIDA, el drástico descenso del número de células *T*4 deja en predominancia las células *T*8 (*derecha*). Las gráficas muestran los resultados de una separación selectiva de células activada por fluorescencia, técnica en la que se utilizan anticuerpos específicos contra las proteínas de la superficie de la célula para marcar las moléculas *T*4 y *T*8 con colorantes que emiten fluorescencias de distintos colores. Las células pasan por un equipo automático que las cuenta y las separa. En las gráficas inferiores, el eje vertical corresponde al número de células y, el horizontal, a la intensidad de la fluorescencia. Las representaciones situadas en el centro combinan las gráficas que relacionan el número de células con la intensidad de la fluorescencia para ambos colorantes. Los máximos sin rotulación corresponden a células que no llevan marcadores *T*4 ni *T*8.

anticuerpos, proteínas que se unen a los antígenos y colaboran en su eliminación o destrucción. Cuando una célula *B* reconoce un antígeno, que puede encontrarse circulando por la sangre o la linfa, o aparece sobre la superficie de una célula infectada, se activa. Se divide entonces para aumentar el clon de células *B* que portan en la membrana moléculas de anticuerpo que actúan de receptores específicos para aquel antígeno. Algunas de las células, las denominadas células plasmáticas, segregan anticuerpo libre. Otras células *B*, de vida más larga, constituyen una raíz de la inmunidad que previene la recurrencia de muchas infecciones. Estas células de memoria permanecen en circulación durante años, siempre dispuestas a crear velozmente una respuesta frente al antígeno si éste vuelve a desafiar al cuerpo.

Las células *T* son más complejas en lo que se refiere a su clasificación y función. En su mayoría no pueden reconocer los antígenos libres en circulación por la sangre o la linfa, y sólo pueden responder a un antígeno situado en la superficie de una célula si se cumplen ciertas condiciones. La sustancia foránea debe exhibirse en conjunción con una de las proteínas propias de la célula del hospedador; en concreto, con alguna de las moléculas codificadas por los segmentos de ADN que constituyen el Complejo Mayor de Histocompatibilidad (MHC). Para que llegue a producirse la respuesta inmunitaria, el receptor del antígeno que se encuentra sobre la superficie de una célula *T* debe reconocer simultáneamente al antígeno y a la proteína MHC.

Tras el reconocimiento del antígeno por parte del linfocito *T*, éste desempeña la función que corresponda a su subclase. Solamente una clase de células *T* defiende al organismo de una manera activa: las células *T* citotóxicas, que destruyen las células infectadas, foráneas o malignas, provocando su lisis, esto es, rompiendo su membrana celular. La otra clase de células *T* modula la respuesta inmunitaria secretando proteínas mensajeras, o bien por contacto directo con las células involucradas.

Las células *T* inductoras desencadenan el proceso de maduración de los linfocitos *T*, a partir de formas precursoras, en células funcionalmente distintas. La participación de células *T* coadyuvantes (*helper*) es condición previa para el correcto funcionamiento de otras células *T* y de la mayoría de las células *B*; reconocido un antígeno específico, permiten que los linfocitos *T* citotóxicos destruyan las células portadoras del antígeno y que los linfocitos *B* segreguen el anticuerpo apropiado. La cuarta clase de células *T*, las supresoras, reprimen la respuesta inmunitaria de las células *B* y *T*, determinando de hecho el cese de la acción inmunitaria defensiva varias semanas después de que la indujera alguna infección. Las células *T* coadyuvantes y supresoras interactúan de manera compleja: ejercen efectos opuestos sobre las células citotóxicas y, al reprimir la respuesta inmunitaria, las supresoras también detienen la acción de las coadyuvantes.



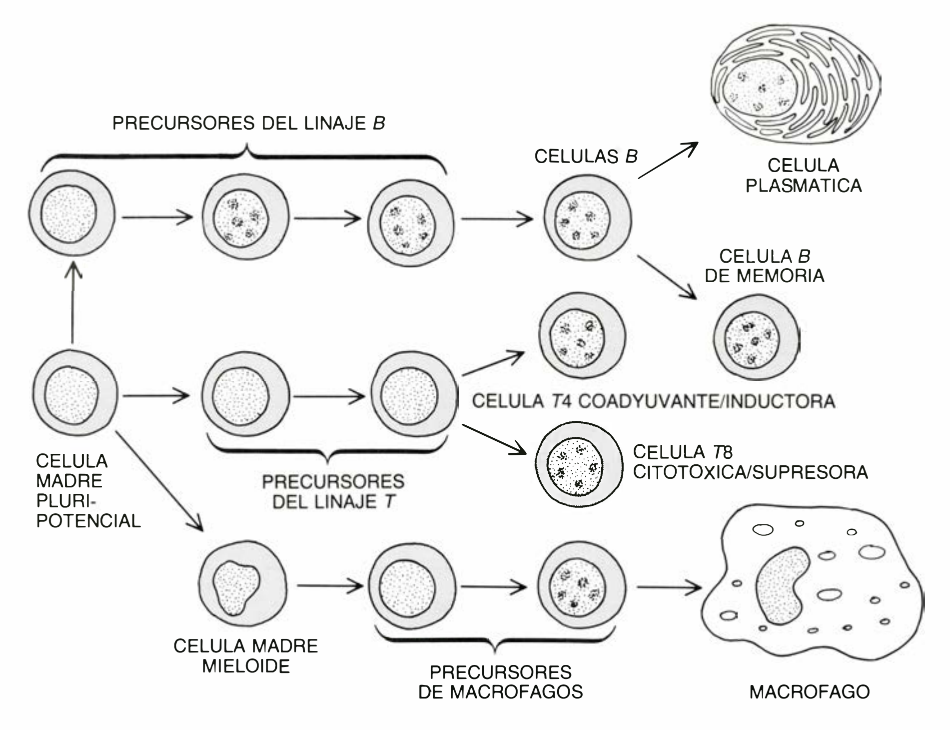
4. SUPERVIVENCIA DE LAS CELULAS *T* en cultivo tras su exposición al virus del SIDA; difiere según el tipo celular de que se parte. El virus afecta poco al número de células *T*8 (*negro*), mientras que rebaja espectacularmente la población de células *T*4, su hospedador preferido (*color*). El virus interfiere la replicación de las células infectadas. Evita también que las células infectadas exhiban el marcador proteico *T*4 sobre la superficie de su membrana; si aún así llegara a exhibirse, una proteína de la cubierta del virus se une a él y lo enmascara. Los dos últimos efectos componen el descenso apreciado en el recuento de células *T*4.

*T*4 y *T*8

Así pues, las células *T* forman cuatro subgrupos en lo que atañe a su función; en lo que se refiere a los marcadores bioquímicos de su superficie, se distinguen, empero, dos clases principales: las células *T*4, que desempeñan papeles de coadyuvantes e inductoras, y las *T*8, con funciones supresoras y citotóxicas. Se distinguen también ambos tipos atendiendo a la clase de proteína MHC que debe ir asociada al antígeno para que puedan reconocerlo. Las *T*8 reconocen el antígeno en el contexto de las proteínas MHC de la clase I, una tipo de moléculas presente en la superficie de todas las células dotadas de núcleo. Consiguientemente, una célula *T*8 citotóxica mata, de ordinario, a cualquier célula infectada que porte un antígeno para el que sea específica la célula *T*8. Las células *T*4, por el contrario, responden a los antígenos asociados a proteínas MHC de la clase II, que se encuentran primordialmente sobre la superficie de las células especializadas conocidas por células presentadoras de antígenos.

Las células presentadoras de antígenos y las asesinas naturales constituyen los otros actores importantes de la respuesta inmunitaria. Entre las primeras, las principales son los macrófagos (células carroñeras que se desarrollan a partir de monocitos, un tipo de glóbulos blancos); se encuentran en la piel y en otros tejidos. Las células de Langerhans, de la piel, y las células dendríticas de la sangre, de los nódulos linfáticos y del bazo, presentan también el antígeno. Los macrófagos actúan englobando un virus o cualquier otro intruso, degradando enzimáticamente sus proteínas, según un procedimiento sumamente específico, y exhibiendo los fragmentos de proteína del antígeno sobre la membrana celular junto con proteínas MHC de la clase II. De este modo, los macrófagos preparan el reconocimiento del antígeno por parte de las células *T*4. Las células asesinas naturales, rama periférica del sistema inmunitario, matan espontáneamente las células infectadas por virus y las células tumorales, sin interactuar directamente con los linfocitos ni reconocer antígenos.

La presencia de un virus invasor



5. DIVERGEN LOS PARTICIPANTES CELULARES en la defensa inmunitaria de sus precursores comunes, las células madre pluripotenciales localizadas en la médula ósea. Las células *B* atraviesan diversos estadios de desarrollo en la médula ósea, tras lo cual se liberan al torrente sanguíneo o la linfa. El contacto con el antígeno las estimula a diferenciarse en células plasmáticas, que segregan anticuerpos, y en células de memoria, responsables de la inmunidad duradera. Los precursores de las células *T* emigran a la glándula del timo para madurar. Después de liberarse a la sangre o a la linfa, las células *T* adquieren distintas identidades bioquímicas y funcionales, así células *T*4, con funciones coadyuvantes o inductoras, o células *T*8, con funciones citotóxicas o supresoras. Los macrófagos comienzan a desarrollarse en la médula ósea como células madre mieloides, que a su vez dan origen a otras clases de glóbulos blancos saguíneos (no mostrados aquí). Los precursores de los macrófagos circulan por la sangre, mientras que los macrófagos maduros emigran a la piel, al bazo y a los nódulos linfáticos, donde es más probable que establezcan contacto con antígenos.

suele desencadenar una compleja interacción entre esos elementos celulares. Para apreciar lo intrincado del fenómeno merecerá la pena reseguir la respuesta inmunitaria frente a una infección vírica ordinaria, por ejemplo, el sarampión. Las células infectadas por el virus secretan ciertas proteínas denominadas interferones, que estimulan una primera línea de defensa contra la infección: las células asesinas naturales. Su actividad aumenta hasta alcanzar un máximo al cabo de un día o dos después de la infección. Entran en liza entonces los macrófagos, que engloban y degradan al virus. Con ello ponen en movimento la respuesta de las células *T*.

Respuestas a la infección

Los macrófagos no sólo estimulan la respuesta inmunitaria alterando los antígenos víricos para que los reconozcan las células *T* coadyuvantes e inductoras, sino también secretando monoquinas, proteínas solubles. Una de las monoquinas secretada por ciertos macrófagos es el interferón gamma; otra, la interleuquina-1, capaz de activar células *T* que hayan reconocido al antígeno vírico sobre la superficie de una célula, preparándolas para diferenciarse y dividirse. La coincidencia de interleuquina-1 e interferón gamma provoca la fiebre y el malestar que acompañan al sarampión y a la mayoría de las enfermedades víricas.

Las propias células *T* activadas producen factores solubles, las denominadas linfoquinas. Precisamente la linfoquina interleuquina-2, secretada por células *T*4 y *T*8, media en el siguiente paso de la respuesta inmunitaria. Bajo la influencia de la interleuquina-2, las células *T* estimuladas por el antígeno y por la interleuquina-1 proliferan en clones mayores de células maduras: células *T* citotóxicas, supresoras y colaboradoras. En una respuesta que alcanza su máximo aproximadamente una semana después de la infección, las células *T* citotóxicas lisan las células infectadas con el virus del sarampión; la eliminación de las células que estén infectadas en ese momento puede detener rápidamente la infección.

Las células supresoras, cuyo número aumenta más despacio que el de células citotóxicas, contribuye a que la respuesta de las células *T* cese al cabo de varias semanas. Tras la supresión persiste una población de células *T* de memoria, que probablemente se mantenga de por vida. Las células *T* de memoria, como las *B* de memoria, intervienen en reacciones de recuerdo frente a posteriores encuentros con el antígeno.

De importancia capital en esa actividad de las células *T* son las células *T*4 coadyuvantes e inductoras. Sin su influencia, ejercida a través de linfoquinas o por contacto directo, no funcionarían las células citotóxicas ni las supresoras. Las *T*4 influyen también en otros aspectos de la respuesta inmunitaria. La interleuquina-2 que segregan mantiene las células asesinas naturales, y producen interferón gamma (considerado una linfoquina cuando lo secretan las células *T*), que estimula a los macrófagos en su misión de englobar el virus y presentar el antígeno.

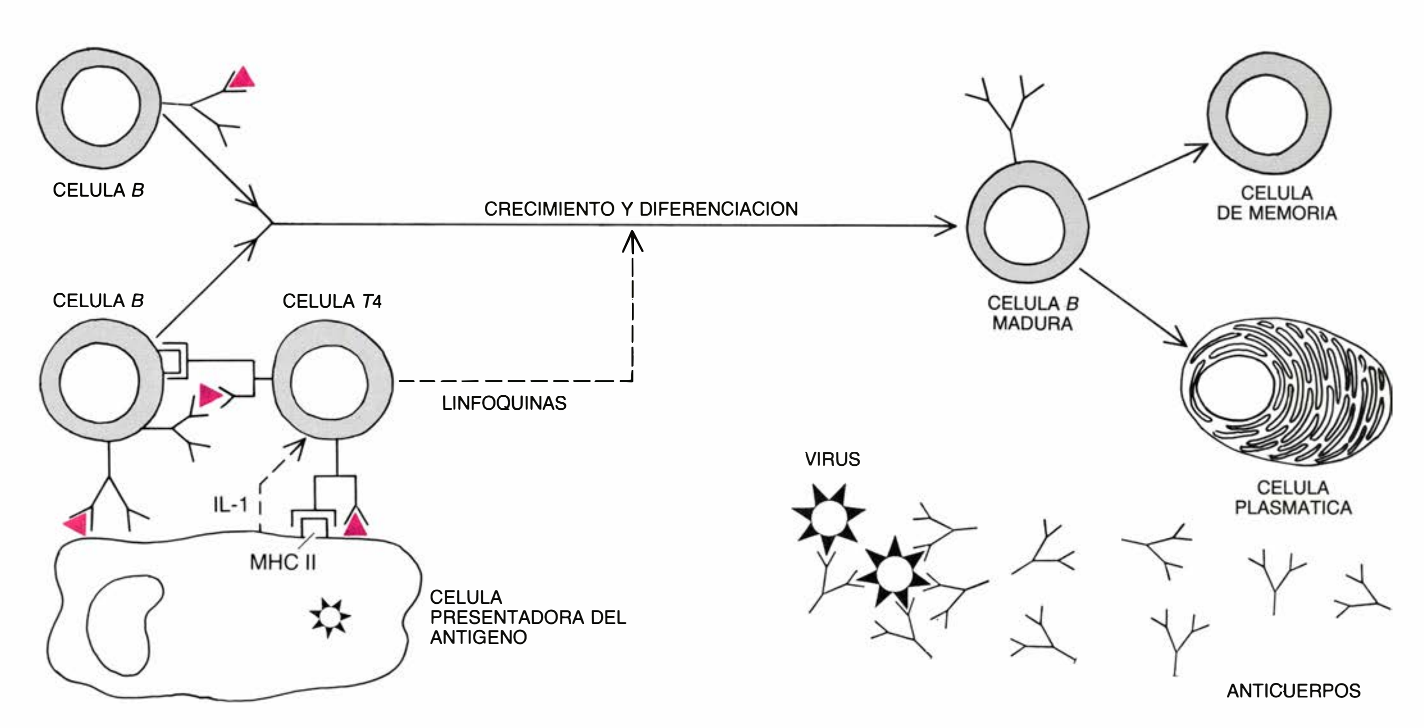
Las células *T*4 resultan igualmente decisivas en la producción de anticuerpos, la segunda y menos importante fase de la defensa del organismo contra una infección vírica, como el sarampión. (Los anticuerpos desempeñan un papel mayor en el fenómeno de la inmunidad duradera.) Para responder al antígeno del sarampión, las células *B* requieren una señal por parte de las células *T* coadyuvantes, bien en forma de linfoquinas o bien por contacto directo. Las células *B* específicas para el antígeno se multiplican luego hasta formar un gran clon de células secretoras de anticuerpo y una población de células *B* de memoria. Las células plasmáticas segregan dos clases principales de anticuerpos contra el virus del sarampión; sus niveles alcanzan el máximo una y tres semanas, respectivamente, después de la infección, antes de que las células *T*8 supresoras (también ayudadas por células *T*4) amortigüen la respuesta de las células *B*.

Anulación de la inmunidad

De no ser por la contraestrategia diseñada por el virus del SIDA, su infección quedaría bloqueada por los mismos mecanismos que vencen al sarampión. No es exclusiva del virus del SIDA la capacidad de evitar la destrucción por parte del sistema inmunitario. Ciertos virus, como el agente de la encefalitis artrítica caprina (CAE), una enfemedad degenerativa de las cabras, lo logran desencadenando una respuesta inmunitaria inapropiada; por ejemplo, propiciando la producción de

ANTIGENO CELULA *T*8
PROLIFERACION
VIRUS
IL-2
CELULA *T*4
AYUDA
SUPRESION
IL-1
MHC II
IL-2
CELULA *T*8
PROLIFERACION
MHC I
CELULA PRESENTADORA DEL ANTIGENO
CELULA INFECTADA

6. COLABORACION DE LAS CELULAS *T* que conduce a la destrucción de una célula infectada por un virus. Una célula presentadora de un antígeno, un macrófago, por ejemplo, ingiere el virus, lo destruye y exhibe las proteínas víricas antigénicas (*color*) sobre su membrana junto con una molécula MHC de la clase II (*MHC II*), una de las propias proteínas del macrófago. Las células *T*4 se activan al unirse simultáneamente al antígeno y a la proteína MHC; la interleuquina-1 (*IL-1*), una proteína secretada por el macrófago, interviene también en la activación de la célula *T*4, que secreta luego interleuquina-2 (*IL-2*). La interleuquina-2 induce la proliferación de las células *T*8 que también han reconocido al antígeno (presentado simultáneamente con una proteína MHC de la clase I). Algunas de las células *T*8 matan a las células infectadas que exhiben el antígeno vírico. Luego, otras células *T*8 se encargan de suprimir esa respuesta citotóxica (*flecha de color*), desconectando la defensa inmunitaria una vez cumplida su misión. Para satisfacer sus cometidos citotóxicos y supresores, las células *T*8 necesitan la colaboración de células *T*4, que se la conceden a través de proteínas solubles o por contacto directo con las células *T*8. (Dibujos de Ilil Abel.)



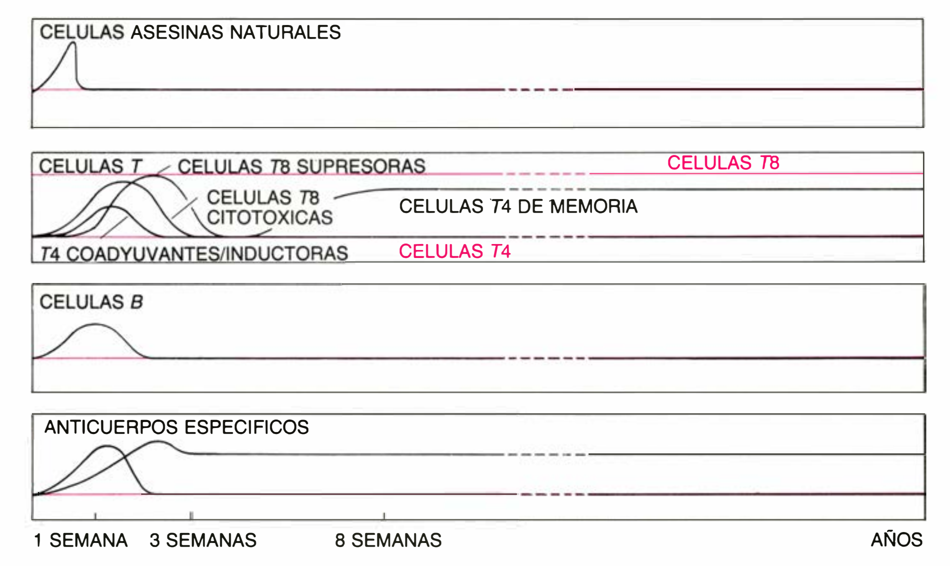
7. LAS CELULAS *T*4 AYUDAN A LAS CELULAS *B* para que secreten anticuerpos contra un antígeno vírico. Una célula *T*4 se activa por medio de la interleuquina-1 (*IL-1*) tras reconocer al antígeno junto con una proteína MHC de la clase II (*MHC II*), ubicadas sobre un macrófago o sobre otra célula de presentación del antígeno. La célula *T*4 se une entonces a una célula *B* que haya reconocido también al antígeno sobre una célula que lo exhiba. El contacto con la célula *T*4 estimula la maduración, multiplicación y diferenciación de la célula *B* en un clon de células de memoria y un clon de células plasmáticas que secretan anticuerpos; éstos se unen al virus, rodeándolo e inactivándolo. Las linfoquinas secretadas por la célula *T*4 colaboran en la maduración. Las células *B* pueden reconocer también antígeno libre en solución en la sangre o en la linfa (*arriba, a la izquierda*), pero no por ello dejan de necesitar la ayuda de las células *T*.

anticuerpos que no bloquean la infección.

Puede también un virus evadir el sistema inmunitario si sus proteínas cambian a menudo por mutación, con lo que cualquier respuesta inmunitaria específica contra un antígeno acabaría perdiendo el rastro. Las ovejas afectas de una enfermedad conocida por visna producen en primer lugar anticuerpos que reaccionan contra el agente vírico, pero el virus muta durante el transcurso de la infección y el anticuerpo pierde toda eficacia. El virus del SIDA, parecido al del visna en algunas de sus proteínas, quizás explote esta variabilidad, que suele denominarse deriva antigénica. Se sabe que el virus sufre constantemente cambios genéticos, debido tal vez a que su replicación es muy rápida y relativamente imprecisa. La deriva antigénica explicaría en parte el hallazgo de que los anticuerpos elaborados por las víctimas del SIDA muestran poco poder neutralizante al someterlos a ensayo frente al virus.

Sin embargo, poca necesidad de evadirse tiene el virus del SIDA; evita su destrucción atacando preferentemente al sistema inmunitario. En otras infecciones víricas se han advertido fenómenos comparables. Incluso el sarampión debilita temporalmente la inmunidad, efecto observado en 1908 por Clemens von Pirquet. En personas que han estado expuestas a la tuberculosis y han adquirido resistencia a la enfermedad, la inyección hipodérmica de una pequeña cantidad de proteínas tuberculosas produce una reacción macroscópica. Von Pirquet observó que los individuos sensibles solían perder la reactividad a la prueba cutánea durante un brote de sarampión, aunque luego la recuperaban.

En contraste con esta pérdida temporal de reactividad, ciertos virus que infectan a los animales producen una depresión duradera de la respuesta inmunitaria. Muchos de ellos son retrovirus, grupo que toma su nombre de una característica única del ciclo biológico de sus miembros. En la secuencia biológica habitual, la información genética, transportada en el ADN, se transcribe a ARN que, seguidamente, se traduce a proteínas necesarias para sostener la vida. Cuando los retrovirus infectan una célula invierten el orden: su información genética, transportada en ARN, se "retrotranscribe" a ADN. En algunos casos, el ADN se integra en los cromosomas de la célula del hospedador, en forma de una secuencia conocida como provirus. Posteriormente la célula hospedadora transcribe los genes víricos y sintetiza las proteínas determinadas por ellos, que se ensamblan formando nuevos virus. Los retrovirus pueden ejercer un profundo efecto sobre el carácter genético, y por tanto sobre las propiedades, de las células que infectan: se sabe que muchos producen cáncer. El conocimiento de la inmunosupresión duradera provocada por algunos retrovirus animales llevó a la suposición, confirmada por el descubrimiento del LAV, es decir, del HTLV-III, de que la causa del SIDA era un retrovirus.



8. SUPRESION INMUNITARIA. Queda patente al comparar la respuesta a una infección típica de sarampión de un paciente con SIDA (*color*) con la respuesta de un individuo con un sistema inmunitario intacto (*negro*). Las curvas relacionan la actividad de las células o de los anticuerpos en función del tiempo. En un individuo normal las células asesinas naturales constituyen la respuesta más temprana; su actividad alcanza el máximo al cabo de uno o dos días. La siguiente respuesta corresponde a las células *T*; sus variedades (*T*4 coadyuvante/inductora, *T*8 citotóxica y *T*8 supresora) proliferan en tiempos distintos, siendo las *T*8 supresoras las últimas que alcanzan el máximo. Tras suprimirse la respuesta de las células *T*, persiste una población de células *T* de memoria que confiere inmunidad frente a una recurrencia de la infección. Mientras tanto, las células *B* se multiplican y secretan anticuerpos, cuyas diversas clases alcanzan sus niveles más altos entre una y tres semanas después de la infección. La supresión anula la actividad de las células *B*, persistiendo una pequeña población de células *B* de memoria, que mantienen altos niveles de anticuerpos durante años. Las células *T*8 de los pacientes de SIDA sobrepasan crónicamente a las células *T*4. Debido al escaso número de células *T*4, y a la alteración de sus funciones, no se producen respuestas inmunitarias específicas.

Agotamiento de células *T*4

En el caso del retrovirus del SIDA, la inmunosupresión responde a la infección vírica de los linfocitos *T*4 que, en el desempeño de su papel de células inductoras y coadyuvantes, orquestan gran parte de la respuesta inmunitaria. La pérdida de tales células en la sangre, nódulos linfáticos, bazo y otros tejidos en los que normalmente se hallan concentradas, constituye uno de los hallazgos más sorprendentes y consistentes en los pacientes de SIDA. De ordinario, las células *T*4 constituyen del 60 al 80 por ciento de la población de células *T* circulantes; con el SIDA llegan a escasear de tal modo que no se detectan. Muchos virus matan las células que infectan, y habitualmente lo hacen desgarrando la membrana celular. Por el contrario, parece que, en cultivo, el virus del SIDA altera, y en última instancia frena, el crecimiento de las células *T*4 infectadas, mientras otras clases de células *T* se multiplican normalmente. Con el tiempo, las células *T*4 desaparecen selectivamente, aunque puede conservarse un pequeño número de ellas, que albergan el virus en estado latente. En trabajos recientes de Robin Weiss, del Instituto de Investigación del Cáncer, de Londres, se ha observado que en las células supervivientes el virus puede enmascarar el marcador *T*4 de la superficie celular, o bien evitar su exhibición. La disminución de células *T*4 parece, por tanto, más espectacular de lo que realmente es.

Las consecuencias de la reducción de la población de células *T*4 reflejan el puesto capital de éstas en el sistema inmunitario. Al faltar la ayuda de las células *T*4, las células *B* son incapaces de producir las cantidades adecuadas de anticuerpo específico contra el virus del SIDA, o contra cualquier otra infección. De manera semejante queda impedida la respuesta de las células *T* citotóxicas; tampoco las supresoras pueden cumplir su papel. Las células *B* de los pacientes con SIDA segregan sin

cesar grandes cantidades de inmunoglobulina inespecífica (la clase de proteínas a la que pertenecen los anticuerpos); no reciben la señal de las células *T* que de ordinario frenaría su actividad.

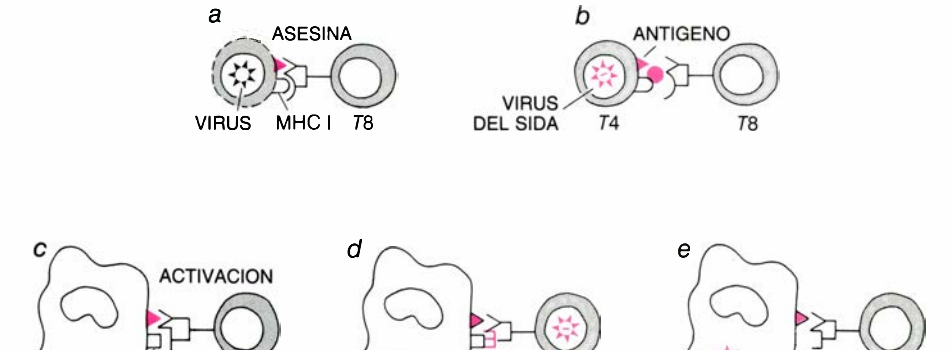
Con la pérdida de las células *T*4 cae el nivel de interleuquina-2, con lo que se frena la expansión clonal de células *T* maduras, normalmente inducida por la linfoquina. La menguada producción de interleuquina-2 y de interferón gamma rebaja la actividad de las células asesinas naturales y de los macrófagos, a los que normalmente estimulan estas proteínas.

La propensión del virus del SIDA a infectar una sola clase de células lo distingue de los demás retrovirus, que tienden a afectar a toda una gama de células. En trabajos recientes se explica esa especificidad. Weiss e, independientemente, un grupo del Instituto Pasteur encabezado por David Klatzmann, han demostrado que una región de la membrana celular asociada al marcador *T*4, la proteína que distingue las células *T*4 de los otros linfocitos, actúa de receptor del virus. Esa región sirve de punto inicial de fijación del virus cuando infecta el linfocito.

Empero, la preferencia del virus por la célula *T*4 no es absoluta; probablemente los macrófagos, las plaquetas y las células *B* actúen de reservorios del virus. La infección de las células *B*, por ejemplo, explicaría su constante secreción de inmunoglobulina. También se acumularía el virus en células no sanguíneas: células endoteliales que tapizan el interior de los vasos sanguíneos y linfáticos, células del epitelio (piel y tejidos emparentados), células de la glía del sistema nervioso y las propias neuronas. La capacidad del virus del SIDA para infectar el sistema nervioso central quizás explique la psicosis y atrofia cerebral que es común en los pacientes. Puede que las células situadas fuera de la sangre carezcan de las proteínas de superficie que permitirían al virus invadirlas directamente, pero se infectarían cuando las células *T*4 o macrófagos enfermos se fusionaran con ellas.

Las células *T*4 quizá resulten especialmente susceptibles a la infección tras haber sido estimuladas, y haber aumentado en número, por parasitosis crónicas o infecciones víricas. Es común entre varios de los grupos de riesgo para el SIDA la infección por el virus de la hepatitis *B*, por el de Epstein-Barr o por citomegalovirus.

La disminución del número de células *T*4 no explica la totalidad de los defectos inmunitarios observados en los pacientes de SIDA. En los primeros estadios de la enfermedad, por ejemplo, los pacientes pueden presentar un número normal de células *T* y, sin embargo, encontrarse sus defensas inmunitarias gravemente debilitadas. Algunos investigadores han propuesto que el virus propicia la producción de anticuerpos contra las células *T*4, los cuales, además de matar a las células *T*4, inhibirían también las células *T*4 supervivientes. Otros sugieren que las células ya infectadas por el virus de la hepatitis *B*, y que portan los genes del virus en su ADN, no sólo resultarían más susceptibles a la infección por el virus del SIDA, sino que responderían también de manera diferente a la infección.



9. DESBARATAMIENTO DEL RECONOCIMIENTO DEL ANTIGENO. Podría constituir uno de los procedimientos por los que el virus del SIDA engaña y confunde al sistema inmunitario. Las células *T* deben reconocer al antígeno en conjunción con una proteína MHC. En circunstancias normales, las células *T*8 citotóxicas destruyen a las infectadas por el virus después de reconocer el antígeno vírico y la molécula MHC de la clase I sobre la superficie de la célula (*a*). Puede que el virus del SIDA obstruya ese proceso determinando una clase alterada de proteína MHC de la clase I (*b*), con lo que evitaría el reconocimiento y destrucción de las células que infecta (en su mayoría células *T*4). También incapacitaría a la propia célula *T*4 para el reconocimiento del antígeno. Las células *T*4 deben unirse al antígeno y a una molécula MHC de la clase II, exhibidos ambos sobre un macrófago (*c*), para que se induzca la proliferación de células *T*. El virus podría destruir el receptor de la célula *T*4 para la proteína MHC (*d*) provocando el fallo del reconocimiento. Podría también incapacitar a las células *T*4 por infección de los macrófagos, obligándoles a exhibir cantidades reducidas de proteínas MHC de la clase II, de tal forma que las células *T*4 no pudiesen unirse al antígeno (*e*).

Factor supresor soluble

El paralelismo entre el virus del SIDA y otros retrovirus inmunosupresores apunta la existencia de algún otro cambio en la función de las células *T*4. Los gatos infectados por el virus de la leucemia felina (FeLV), un retrovirus, a menudo no mueren de leucemia, sino de otras afecciones, entre las que figuran algunas de las enfermedades infecciosas oportunistas encontradas en el SIDA. En 1978, Lawrence E. Mathes y Richard G. Olsen, de la Facultad de Veterinaria de la Universidad estatal de Ohio, descubrieron un fundamento de la supresión inmunitaria que explica la susceptibilidad de los gatos ante tales enfermedades.

Mathes y Olsen advirtieron que una parte de la cubierta proteica del FeLV inhibe la respuesta, de las células *T* de gato cultivadas, frente a una sustancia que, de ordinario, en una respuesta inmunitaria normal, insta la proliferación celular. Los investigadores comprobaron que la molécula de la cubierta, que se proyecta desde la superficie de las células infectadas por FeLV, altera la respuesta inmunitaria en el animal vivo. Tras inyectar en gatos la proteína purificada, denominada p15E, encontraron que, en esos animales, la probabilidad de desarrollar cáncer después de una exposición al virus del sarcoma felino era superior a la de los gatos normales. Desde entonces, cada vez son más las pruebas, reunidas por diversos laboratorios, de que otros retrovirus de mamíferos poseen componentes proteicos que producen deficiencias inmunitarias.

Tales deficiencias inmunitarias debidas a los retrovirus comparten muchas de las características del SIDA, por ejemplo, el desbaratamiento de todas las respuestas que dependen de células *T*4: la producción de anticuerpos, la actividad citotóxica de las células *T*, la proliferación de células *T* inducida por la interleuquina-2 y la estimulación de los macrófagos por el interferón gamma. Henry G. Kunkel y el autor, trabajando en la Universidad Rockefeller, avanzaron la hipótesis de que en

el SIDA podría estar implicado un factor proteico. En 1982 demostraron que las células sanguíneas cultivadas, procedentes de pacientes que padecían el SIDA, liberaban un factor que puede inhibir ciertas respuestas inmunitarias. La sustancia, que denominaron factor supresor soluble, bloqueaba las respuestas inmunitarias dependientes de células *T*, así la producción de anticuerpo específico y la proliferación de células *T*, tanto in vitro como al inyectarla en ratones.

Demostraron a continuación que la fuente de factor supresor soluble eran las células *T*4 infectadas. Se observaron células *T*4 de un paciente que elaboraba grandes cantidades del factor y se fusionaron con células *T* normales para inmortalizarlas. Se observó que los clones de células híbridas resultantes producían el factor soluble en abundancia. Más tarde, trabajando con Philip D. Markham, en el laboratorio de Gallo, expusieron células *T*4 normales al virus del SIDA, comprobando que la infección desencadenaba la secreción de un factor supresor.

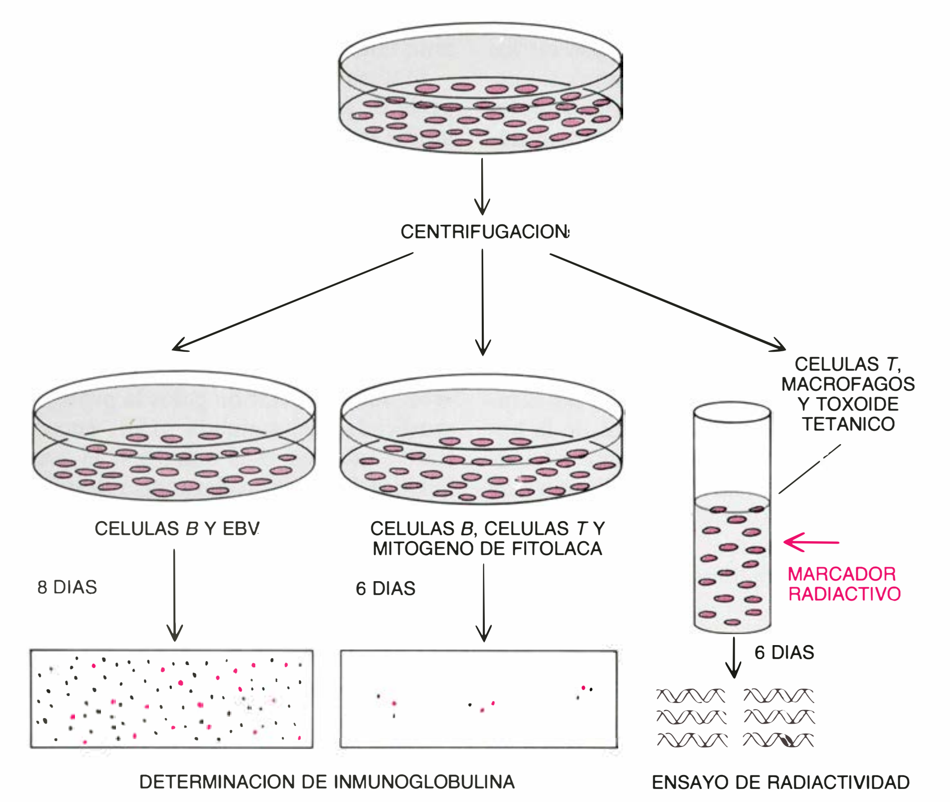
Con el ánimo de determinar si ese factor supresor soluble estaba relacionado con el p15E, George J. Cianciolo y Ralph Snyderman, del Hospital Clínico de la Universidad de Duke, compararon las secuencias de aminoácidos que constituyen las proteínas de la cubierta del virus del SIDA con las de p15E. En la comparación incluyeron además otro retrovirus con afinidad por las células *T* humanas: el HTLV-I. Un fragmento de 26 aminoácidos de una proteína de la cubierta del HTLV-I coincidía con p15E en el 73 por ciento de los aminoácidos. También una proteína de la cubierta del virus del SIDA guardaba semejanza, en este caso menor, pero aún considerable, con una región de p15E. Se planteó, por tanto, la hipótesis de que el factor supresor soluble procediera de la cubierta del virus del SIDA.

Malcom A. Martin y sus colaboradores, del Instituto Nacional de la Alergia y Enfermedades Infecciosas de los Estados Unidos, proponen un origen distinto del factor. Observaron que la región del genoma de FeLV que determina p15E y la región del genoma de HTLV-I que codifica la proteína de la cubierta que encaja con aquella guardan semejanza con un segmento de ADN humano. Se cree que ese segmento corresponde al código genético de un retrovirus endógeno: un retrovirus que, tiempo atrás, se integró en la dotación genética humana. Quizás el virus del SIDA haga que la célula hospedadora exprese la secuencia endógena (es decir, que fabrique la proteína que codifica). Más que determinar el factor supresor soluble en su gen de la cubierta, el virus del SIDA simplemente induciría la síntesis del factor por parte de la célula *T*4 hospedadora. Ese mismo mecanismo podría subyacer a la inmunosupresión observada en otras infecciones.

Según han demostrado Cianciolo y Snyderman, cualquiera de las dos explicaciones resulta plausible. Sintetizaron un péptido (una breve cadena de aminoácidos) que contenía una serie de aminoácidos comunes a p15E, a las secuencias de la cubierta de HTLV-I y al virus del SIDA, así como la proteína codificada por la secuencia retrovírica humana. El péptido, de 17 aminoácidos, era capaz, según encontraron los investigadores, de inhibir las mismas respuestas que la molécula intacta de p15E, que consta de 196 aminoácidos. El resultado sugiere intensamente que todas las secuencias ejercen efectos inmunosupresores.

Además de reducir el número de células *T*4 y provocar la liberación de un factor supresor soluble por parte de las supervivientes, el virus del SIDA impide a las células *T*4 que sobreviven desempeñar el primer y decisivo paso de la respuesta inmunitaria: el reconocimiento del antígeno. Anthony S. Fauci y H. Clifford Lane, del Instituto Nacional de la Alergia y Enfermedades Infecciosas, han demostrado, por ejemplo, que las células *T* de pacientes con SIDA no responden a un antígeno bacteriano común, como es el toxoide tetánico. Cuando esos investigadores expusieron células *T* al antígeno, en presencia de macrófagos (imprescindibles para que se elabore y presente el antígeno), las células no proliferaron, mientras que sí lo hubieran hecho las células *T* normales.

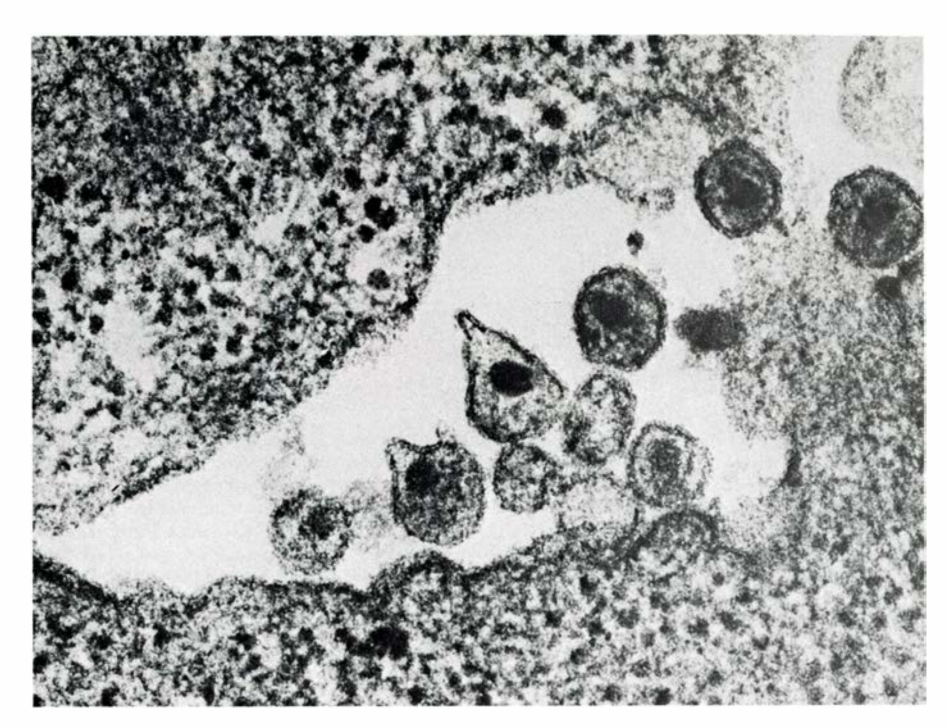
Una posible explicación de ello es que el virus altere, de alguna manera, el receptor para el antígeno situado en



10. FACTOR SUPRESOR SOLUBLE producido por linfocitos *T*4 infectados de pacientes de SIDA, o del complejo relacionado con el SIDA (ARC); inhibe la respuesta inmunitaria que depende de células *T*. En el laboratorio del autor se cultivaron glóbulos blancos procedentes de pacientes de SIDA (*arriba*), sometiéndose luego a centrifugación. Se añadió el sobrenadante, o líquido libre de células, a tres cultivos de células sanas. Uno contenía células *B* y virus de Epstein-Barr (EBV), que de ordinario induce a las células *B* a proliferar y a secretar inmunoglobulina sin ayuda por parte de las células *T*. La respuesta de las células *B* al EBV no se vio afectada. El segundo contenía células *B*, células *T* y el mitógeno fitolaca, un estímulo ante el cual las células *B* sólo pueden secretar anticuerpo con ayuda de las células *T*. Las células *B* secretaron aquí poco anticuerpo. El tercer cultivo contenía células *T*, macrófagos (necesarios para exhibir el antígeno ante las células *T*) y toxoide tetánico. En circunstacias normales, las células *T* previamente sensibilizadas proliferan en respuesta al antígeno del tétanos. La adición al cultivo de un nucleótido (una subunidad de ADN) marcado radiactivamente, seguida de una valoración de la cantidad de radiactividad incorporada al nuevo ADN, indicó que las células *T* no se habían multiplicado. Algún factor del sobrenadante suprimió su respuesta.

la superficie de las células *T*4 (la cerradura molecular en la que debe encajar la llave constituida por el antígeno y una proteína MHC de la clase II para que se desencadene la respuesta de la célula *T*). El virus codificaría una proteína expresada sobre la superficie de la célula infectada, que invadiría y destruiría el mecanismo receptor. Se sabe que ciertos retrovirus que infectan ratones desbaratan los receptores situados sobre los linfocitos *T*, por lo que las células *T* no pueden distinguir ningún antígeno. Sin embargo, a diferencia de las células *T* incapacitadas de las víctimas del SIDA, los linfocitos de ratón matan indiscriminadamente las células infectadas y las sanas.

Un mecanismo similar podría proteger a las propias células infectadas del SIDA para que no las reconociera el sistema inmunitario. Un virus cancerígeno denominado Ad-12 bloquea la transcripción de ciertas secuencias genéticas en las células infectadas, haciendo que las células expresen sobre su superficie cantidades reducidas de proteínas MHC de la clase I. Puesto que las células *T*8 citotóxicas sólo pueden unirse al antígeno si está junto con la proteína MHC de la clase I de la propia célula hospedadora, el efecto entorpece el reconocimiento y la destrucción de las células infectadas. Existen algunas pruebas de que el virus del SIDA actúa de manera semejante, protegiendo de ese modo a las células *T*4 y a las demás células que infecta. Este efecto podría ayudar al virus a eludir cualquier vestigio de sistema inmunitario que quedara en activo.



11. PARTICULAS DE VIRUS DEL SIDA, distinguibles por su núcleo oscuro de ARN. Geman de la superfice de células *T* infectadas. La micrografía, tomada en el Instituto Pasteur de París en 1983, corresponde a la primera imagen registrada en el aislamiento de las partículas víricas, que se denominaron LAV; al año siguiente, ese mismo virus se identificó en el Instituto Nacional del Cáncer de los Estados Unidos con otro nombre, HTLV-III. El virus del SIDA es un retrovirus: tras penetrar en una célula *T*4 "retrotranscribe" su dotación genética, incardinada en ARN, a ADN de la célula hospedadora. Posteriormente induce a la célula hospedadora a transcribir los genes retrovíricos y fabricar ARN y proteínas. Los componentes víricos forman nuevos virus, que se liberan de la célula. La ampliación de la micrografía es de 95.000 diámetros.

Tratamiento

¿Dónde podría intervenirse, en este embrollo patológico, para curar la enfermedad o reducir sus efectos? La pregunta pide una respuesta urgente, y no sólo por el número de casos ya diagnosticados, sino también porque el SIDA, según su definición actual, probablemente se atribuya sólo a una pequeña proporción de las personas infectadas por el virus. Durante los últimos años se advierte cada vez con mayor frecuencia un agrandamiento persistente y generalizado de los nódulos linfáticos entre miembros, por lo demás sanos, de los grupos de alto riesgo para el SIDA. El examen microscópico de los nódulos, después de su extirpación por métodos quirúrgicos, a menudo pone de manifiesto una desaparición de células *T*4 y el desbaratamiento de la arquitectura celular, observados también en los casos bien manifiestos de SIDA. Se cree que muchas de esas personas desarrollarán en primer lugar el complejo relacionado con el SIDA (ARC), caracterizado por una fiebre sin explicación, sudor nocturno, pérdida de peso, tos crónica o diarrea, y luego el propio SIDA.

Los análisis de sangre sugieren que es aún mayor el número de personas posibles portadoras de la infección, pese a no mostrar síntoma alguno de enfermedad. La presencia en la sangre de anticuerpos contra el virus del SIDA indica que ha habido una exposición anterior al virus y, en muchos casos, pone de manifiesto la infección: la mayoría de los varones homosexuales con anticuerpos tienen el virus infeccioso en su sangre y en algunos casos también en el semen, saliva y lágrimas. La presencia de anticuerpos entre personas con riesgo de contraer la enfermedad permite calcular que, sólo en Estados Unidos, entre uno y dos millones de personas están infectadas por el virus del SIDA. Algunas de ellas quizá no muestren jamás los síntomas, aun cuando las infecciones retrovíricas persistan durante toda su vida. Para otras, el período de incubación varía entre meses y décadas. Según un cálculo, probablemente conservador, el 7 por ciento de los individuos actualmente infectados, pero todavía sanos, desarrollará el SIDA cada año.

Igualmente preocupante es que esas personas constituyen un gran depósito de portadores capaces de propagar la enfermedad. Dado que una pequeña proporción de los portadores del SIDA no producen cantidades detectables de anticuerpos contra el virus, los procedimientos aplicados hoy para examinar la sangre de donantes no permiten obviar por completo el riesgo de infección por transfusión sanguínea.

Al tratar los casos activos de SIDA, los médicos suelen concentrarse en las manifestaciones clínicas de la enfermedad: las infecciones y cánceres que se desarrollan a causa de su deficiencia inmunitaria. El éxito a largo plazo es raro: casi todos los pacientes con SIDA a los que se les ha diagnosticado una infección oportunista mueren en el plazo de cuatro años. Otras estrategias de tratamiento intentan restaurar en parte la función inmunitaria. Se han ensayado transplantes de médula ósea e inyecciones de glóbulos blancos para reponer células sanas en el sistema inmunitario, al que se ha intentado estimular administrando interleuquina-2 e interferones. De momento, tampoco esos esfuerzos se han visto premiados.

	NOMBRE DE LA PROTEINA	SECUENCIA DE AMINOACIDOS
VIRUS DEL SIDA	gp41	L Q A R I L A V E R Y L K D Q Q L
HTLV-1	gp21	Q N R R G L D L L F W E Q G G L
FeLV	p15E	Q N R R G L D I L F L Q E G G L
RETROVIRUS 4-1 ENDOGENO HUMANO	p15E	Q N R L A L D Y L L A A E G G V
PEPTIDO SINTETICO	CKS-17	L Q N R R G L D L L F L K E G G L

12. SECUENCIAS DE AMINOACIDOS de las proteínas retrovíricas que pueden suprimir el sistema inmunitario, entre las que se advierten estrechas semejanzas. Cada letra representa uno de los 20 aminoácidos que forman las proteínas; las sombras grises y de color agrupan aminoácidos con propiedades similares. Utilizando una proteína inmunosupresora conocida procedente de la cubierta proteica del virus de la leucemia felina (FeLV), George J. Ciancolo y Ralph Snyderman, del Hospital Clínico de la Universidad de Duke, identificaron secuencias similares en las proteínas de la cubierta del virus del SIDA y del HTLV-I (un virus que infecta células *T* humanas) y en una proteína codificada por un segmento del genoma humano. Se cree que el segmento es un retrovirus endógeno. Los investigadores sintetizaron luego un péptido que contenía aminoácidos compartidos por las secuencias. El péptido sintético suprimía las respuestas inmunitarias con la misma eficacia que la proteína de FeLV, señal de que las otras proteínas probablemente también sean inmunosupresoras. Cabe suponer, por tanto, que el virus del SIDA podría producir una proteína supresora en su cubierta, o bien inducir a la célula a sintetizar la proteína codificada por esa secuencia endógena.

Otra manera de proceder es la búsqueda de un arma contra el propio virus. Investigadores del laboratorio de Gallo han descubierto que la droga suramina, utilizada hasta ahora para tratar infecciones por protozoos, es capaz, in vitro, de evitar que el virus del SIDA infecte y dañe las células *T*. El cuerpo humano aceptaría la concentración necesaria de la droga. La suramina actúa inhibiendo la enzima del retrovirus, la transcriptasa inversa, que participa en la transcripción del ARN vírico a ADN del hospedador (el primer paso de la infección). Resultarían de pareja utilidad otras drogas que actúan en contra de la transcriptasa inversa o que interfieren en pasos posteriores del ciclo biológico del virus, como son la producción de proteínas de acuerdo con los genes del virus y el ensamblaje de nuevos virus. Entre las sustancias propuestas o ensayadas se cuentan la ribavirina, anteriormente utilizada contra los virus del catarro común y de la gripe; un nuevo compuesto denominado HPA-23 y el fosfonoformato; la 3′-azido-3′-desoxitimidina, un miembro de la familia de agentes anticancerosos; la ansamicina, emparentada con una droga que se administra para la tuberculosis, y el interferón alfa. Desgraciadamente, muchos de los fármacos que inhiben la transcripción inversa o la replicación del virus alteran también el crecimiento de las propias células del hospedador, incluidas, irónicamente, las del sistema inmunitario.

En mi laboratorio de Cornell, mis colegas y yo estamos investigando sobre drogas antivíricas que, por añadidura, son capaces de estimular el sistema inmunitario del hospedador. Los compuestos están emparentados con la guanosina, uno de los bloques estructurales del ADN y el ARN; uno de los fundamentos de nuestro interés por esas sustancias es que una de ellas, el aciclovir, ha probado su eficacia contra otra infección vírica, la del herpes simplex. In vitro, esas sustancias no inhiben la replicación del virus del SIDA, y además incrementan la respuesta de los linfocitos de los pacientes con SIDA frente a compuestos que inducen la proliferación de células *T*. Las drogas, que al parecer actúan inhibiendo enzimas fundamentales para la síntesis de ARN y ADN víricos, son prometedoras, y destacan la necesidad de investigar las interacciones de los fármacos antivíricos con el sistema inmunitario.

¿Una vacuna contra el SIDA?

Para el creciente número de infectados, la esperanza se centra en dar con una droga antivírica eficaz. La necesidad de una vacuna contra el SIDA no es menos acuciante. La variabilidad genética del virus entorpecerá la búsqueda de una vacuna; las muestras del virus aisladas a partir de diversos pacientes difieren hasta en más del 30 por ciento en las secuencias de ARN que codifican las proteínas supuestamente fundamentales para el reconocimiento por parte de las células *T* y los anticuerpos. Las vacunas estimulan el sistema inmunitario con un antígeno, propiciando la producción de anticuerpos y la proliferación de células de memoria. La variabilidad del virus del SIDA supondría que una posterior exposición al virus quizá no despertara la memoria inmunológica creada por la vacuna. Sin embargo, cabe aún identificar regiones invariables de la cubierta vírica, a las que pudiera unirse eficazmente el anticuerpo, y utilizar esas regiones para la base de una vacuna. El reciente avance logrado por Olsen y sus colaboradores en la preparación de la primera vacuna activa contra un retrovirus de mamífero, el FeLV, ha animado a otros investigadores.

Los trabajos de William Haseltine y Joseph G. Sodroski, en la Escuela de Sanidad Pública de Harvard, apuntan un posible enfoque de la vacuna contra el SIDA. Han comprobado esos autores que las células *T* infectadas por HTLV-I o por el virus emparentado HTLV-II secretan un factor regulador que incrementa la transcripción del propio virus y de secuencias ubicadas en el ADN de las células del hospedador. También han visto que el genoma del virus del SIDA contiene una secuencia, conocida por *tat*, que codifica una proteína reguladora similar. En el caso del virus del SIDA, la proteína podría estimular la transcripción de genes víricos (y tal vez del gen vírico o del hospedador para el factor supresor soluble), bien inhibiendo genes que estimulan la replicación de la célula *T*4 hospedadora, o bien activando genes que detienen la división celular. Además de localizar la región del ARN que codifica la proteína reguladora, Haseltine y Sodroski han identificado la secuencia de ARN con la que interactúa para estimular la transcripción.

Si pudiera modificarse genéticamente el virus del SIDA por deleción de *tat* o de la secuencia con la que interactúa la proteína de *tat*, serviría para preparar una vacuna segura y fiable. Propiciaría una respuesta inmunitaria que bloquease una posterior infección por el virus sin modificar, pero el virus aberrante no provocaría una infección dispersada ni agotaría las células *T*4. Alternativamente, una droga que inhibiese la síntesis de la proteína reguladora codificada por *tat* constituiría una defensa química, en oposición a la inmunológica, contra la infección por parte del virus del SIDA. Cualquiera de las dos estrategias de protección contra el virus sería una bendición para las decenas de millones de personas que se cuentan hoy entre los grupos de riesgo para el SIDA.

El sistema alcoholdeshidrogenasa

El estudio de este gen en distintas especies de Drosophila *y de sus productos de expresión permite abordar cuestiones esenciales que se plantean hoy en biología molecular; entre éstas: el mantenimiento de variantes y la evolución de los genes*

Roser Gonzàlez-Duarte, Elvira Juan, Lluïsa Vilageliu y Sílvia Atrian

La especie *Drosophila melanogaster*, la mosca del vinagre, constituye un excelente material de trabajo tanto para estudios genéticos y evolutivos como de biología molecular, merced a su fácil cultivo en el laboratorio y a la gran cantidad de conocimientos que se han acumulado a lo largo del tiempo sobre este organismo eucariota. Utilizada a principios de siglo por citólogos y genetistas para someter a prueba, entre otros, los postulados de la teoría cromosómica de la herencia y de la mutación, contribuyó más adelante al esclarecimiento de los procesos de especiación y es hoy la base de muchos estudios de genética molecular y evolutiva, amén de uno de los organismos superiores que se ofrece de modelo para conocer los patrones del desarrollo animal.

Las unidades de la herencia son los genes, que determinan proteínas. En *Drosophila*, el gen alcoholdeshidrogenasa (*Adh*) y su producto de expresión, la enzima ADH, han sido también objeto de múltiples investigaciones, tanto desde una perspectiva genética como bioquímica, para resolver cuestiones esenciales que se plantea hoy la biología molecular: el proceso de adaptación de los organismos al medio en que se encuentran, el mantenimiento de variantes genéticas en poblaciones naturales, la evolución de los genes en distintas especies y los mecanismos de regulación de la expresión génica.

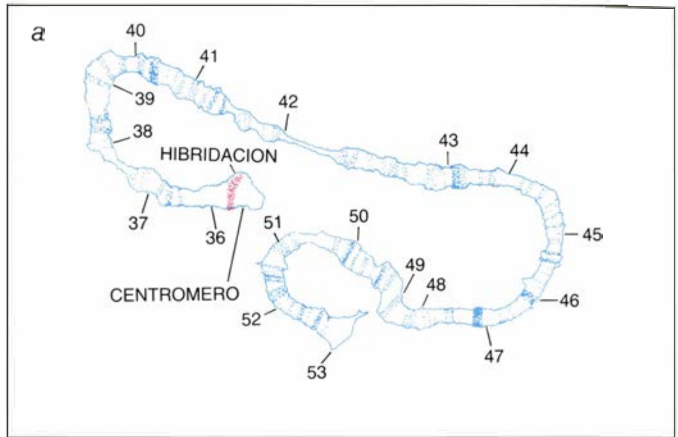
En nuestro departamento de genética de la Universidad de Barcelona, nos propusimos, desde hace varios años, un triple objetivo: analizar la evolución de la enzima ADH en el nivel bioquímico y estructural, estudiar las causas de su posible función adaptativa en el nivel molecular y precisar su función fisiológica en relación con otras enzimas implicadas en el metabolismo del alcohol. Se eligieron siete especies del género *Drosophila: simulans, virilis, lebanonensis, funebris, immigrans, hydei* y *melanogaster*; esta última servía de punto de referencia. Las siete especies, que se distribuyen en tres subgéneros distintos, difieren ampliamente entre sí en lo concerniente a la tolerancia al alcohol, ocupan asimismo distintos nichos ecológicos y presentan una distancia filogenética diferente con respecto a *Drosophila melanogaster*.

De entrada, había que comparar el producto del gen *Adh*, la enzima ADH, en su nivel bioquímico y estructural, en las siete especies. Para ello era necesario purificar la enzima, de modo que pudiéramos disponer de preparaciones libres de los contaminantes más frecuentes, tales como proteasas, que producen una degradación rápida de la proteína pura y restan finura a los resultados experimentales. Por alelo se entiende cada una de las formas alternativas de un gen que se ubica en un lugar concreto (*locus*) del mapa genético. Cada alelo determina la síntesis de una de las formas alternativas de una enzima, denominadas aloenzimas. Se sabe que, en un mismo locus cromosómico de *D. melanogaster*, se sintetizan, por lo menos, tres variantes genéticas de la ADH, aloenzimas. Su síntesis está controlada por tres alelos: Adh^S, Adh^F y Adh^{UF}. Las letras *S, F* y *UF* se han asignado en función de la movilidad electroforética de la ADH en un gel de almidón; este mismo orden refleja la movilidad de la enzima hacia el cátodo, siendo *S* la variante más lenta, *F* la intermedia y *UF* la más rápida en determinadas condiciones establecidas. A partir de una cepa homocigótica para el alelo Adh^S de *D. melanogaster*, la enzima ADH^S había sido aislada y caracterizada por David Thatcher en 1977, a la sazón en la Universidad de Edimburgo.

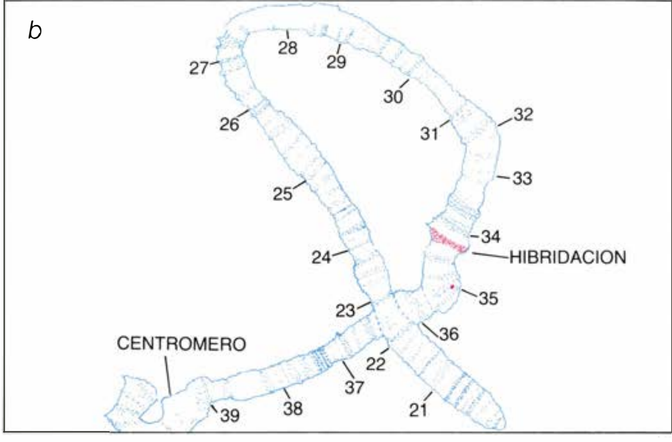
Decidimos utilizar la enzima de esta cepa como referencia y mejorar en lo posible el proceso de purificación descrito, introduciendo una cromatografía de afinidad: técnica que se basa en la afinidad de un compuesto, en nuestro caso una enzima, hacia un determinado producto para separarlo del resto de los componentes de una solución. Nos servimos de una resina de Sepharosa que contenía enlazado a la matriz un análogo del NAD^+, producto que emplean algunas deshidrogenasas como cofactor. Las enzimas ADH de nuestras especies presentaban una gran afinidad al NAD^+ y, por tanto, se enlazarían fuertemente con este producto análogo. Así pues, al introducir en la columna de afinidad una solución semipura de ADH quedaba retenida la enzima, que se eluía al incrementar bruscamente la

1. LOCALIZACION CITOLOGICA del gen *Adh* en *Drosophila* merced a la hibridación de los cromosomas politénicos de las glándulas salivales. Se emplea un plásmido recombinante, *s*AC1, que contiene el gen *Adh* de *D. melanogaster*, marcado con biotina. La detección de los híbridos de ADN se lleva a cabo mediante un complejo de estreptavidina-peroxidasa biotinada. La estreptavidina de este complejo se une a la biotina incorporada en el ADN; la peroxidasa, en presencia del substrato y colorante adecuados, reacciona produciendo un precipitado marrón oscuro. Las señales que se obtienen en los cromosomas sirven de punto de referencia para comparar los modelos de bandas adyacentes a dichas señales, establecer la homología entre los cromosomas de las distintas especies y determinar las reordenaciones que han tenido lugar durante la evolución. El gen *Adh* se localiza en cromosomas homólogos de las especies estudiadas, pero a distinta distancia del centrómero. Ello demuestra que se han producido inversiones de fragmentos cromosómicos, pericéntricas o paracéntricas –según contengan el centrómero o no–, o ambas a la vez. La localización del gen *Adh* en *D. subobscura* (*A*) y *D. madeirensis* (*D*) con relación a *D. melanogaster* (*B*) se explica mediante una inversión paracéntrica ocurrida a lo largo de la evolución. Además, en el grupo *obscura* debió producirse una inversión pericéntrica y otra paracéntrica situando el gen *Adh* de *D. ambigua* (*C*) alejado del centrómero y próximo al extremo del brazo derecho del cromosoma, si bien no puede asignarse un orden temporal a estos sucesos. Se indica las subdivisiones o regiones cromosómicas, para la localización de la *Adh* en las distintas especies, establecidas por diversos autores. Las de *D. madeirensis* nos las facilitó M. Papaceit.

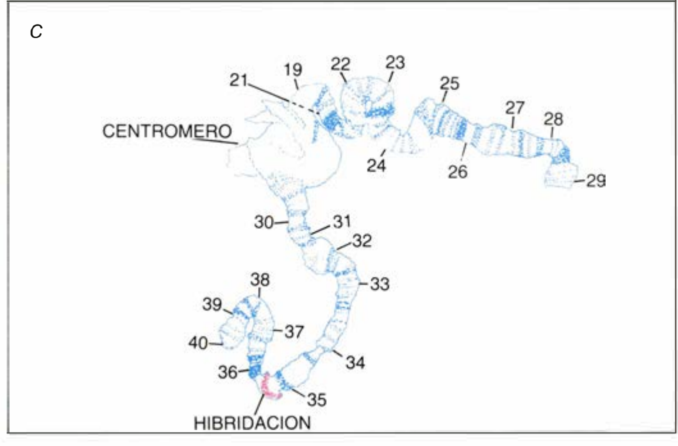



a
40
41
39
42
38
43
44
HIBRIDACION
37
36
51
50
45
CENTROMERO
49
48
46
52
47
53

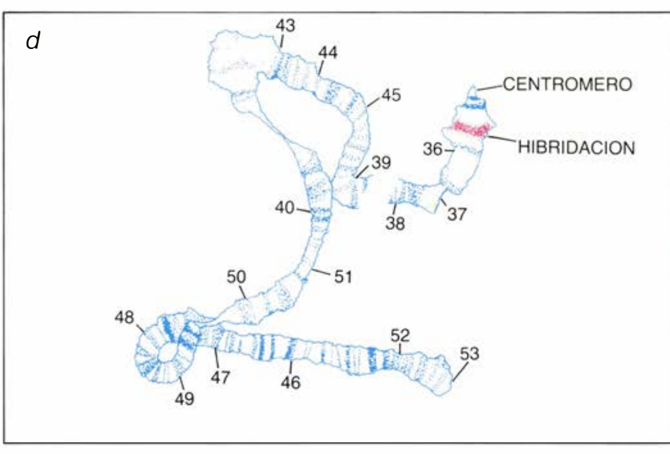


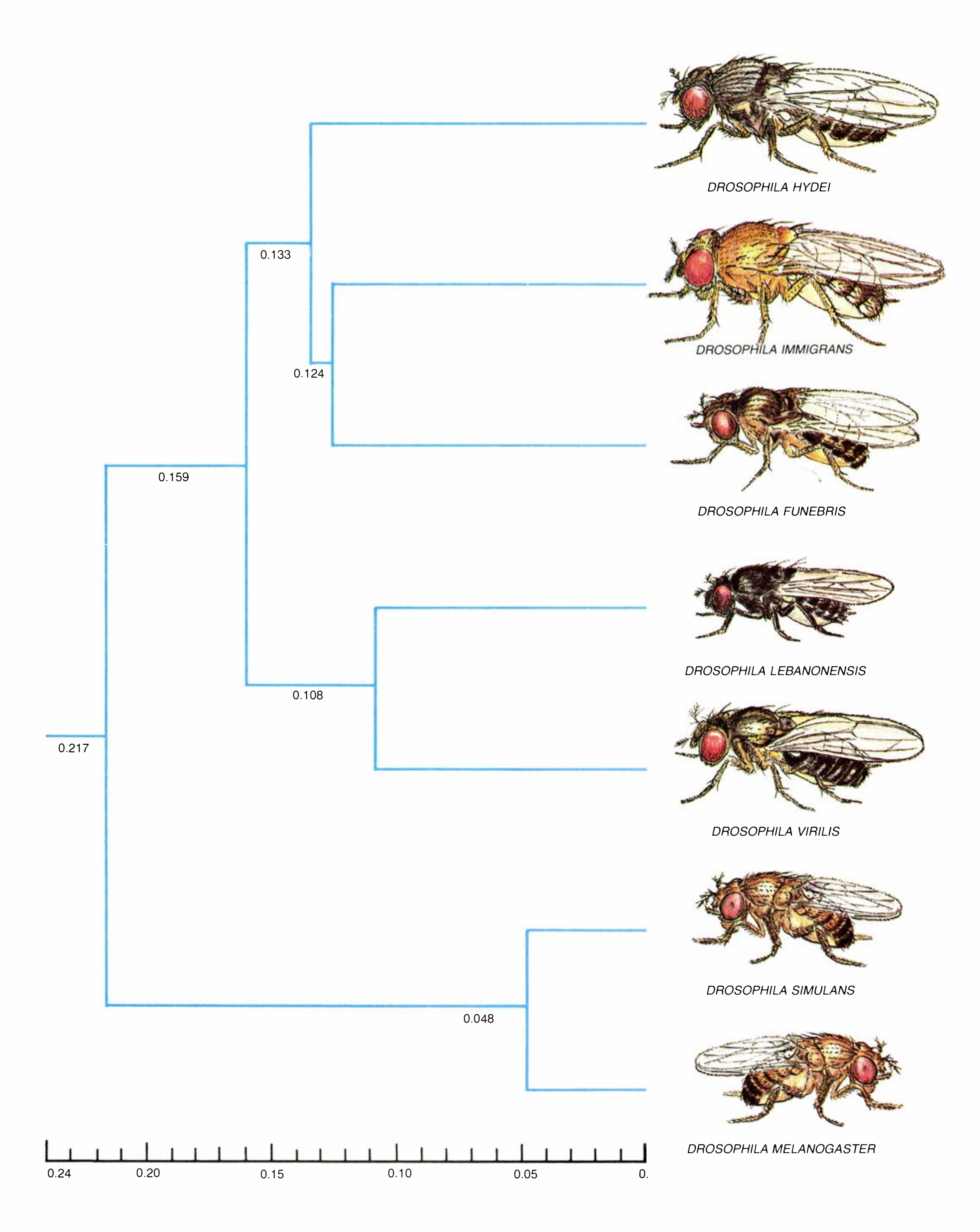

b
28
29
27
30
26
31
32
33
25
34
24
HIBRIDACION
35
23
36
CENTROMERO
22
37
38
39
21




c
19
22
23
21
25
27
28
CENTROMERO
24
26
29
30
31
32
38
33
39
37
40
34
36
35
HIBRIDACION




d
43
44
CENTROMERO
45
HIBRIDACION
36
39
40
37
38
50
51
48
52
53
47
46
49



2. RELACIONES FILOGENETICAS de varias especies de *Drosophila*, basadas en la homología de la ADH. Las distancias genéticas entre las especies se han calculado según el parámetro de Bhattacharyya: $d_{\alpha\beta} = 2 \arccos (\Sigma \sqrt{p_i\, q_i})$ donde $p_1 \ldots, p_n$ son las frecuencias relativas de los aminoácidos de la ADH en la especie α y $q_1, \ldots q_n$ las frecuencias relativas de los aminoácidos en la especie β. (El sumatorio Σ abarca desde $i = 1$ hasta $i = n$.) Una vez calculadas las distancias, valores situados en las ramificaciones, se construyó el dendrograma utilizando el método *UPGMA*. Según este dendrograma, las siete especies aparecen agrupadas en tres ramas: *D. melanogaster* y *D. simulans*; *D. virilis* y *D. lebanonensis; D. immigrans*, *D. funebris* y *D. hydei*. Esta distribución está de acuerdo con las relaciones filogenéticas establecidas en el género *Drosophila* después de analizar un elevado número de rasgos morfológicos y del estudio de los cromosomas.

concentración de NAD^+. Se observaba frecuentemente la presencia de un contaminante, otra deshidrogenasa quizá, que se eliminaba en una cromatografía posterior de filtración en gel. Todas las pruebas que realizamos para verificar la pureza de las preparaciones de la ADH revelaron la ausencia de contaminantes. Ni siquiera a concentraciones muy elevadas de enzima se logró detectar la presencia de otra proteína.

Utilizando preparaciones puras de enzima de las distintas especies iniciamos los estudios de caracterización bioquímica. Sorprendía la estabilidad de la enzima pura de varias especies en solución a 4 grados Celsius. La vida media de la ADH de *D. virilis, D. simulans, D. melanogaster* (Adh^S) y *D. hydei* era igual o superior a tres meses. *D. funebris* y *D. immigrans* presentaban una estabilidad menor, de 45 y 18 días respectivamente. Algunos de los parámetros bioquímicos clásicos que definen las enzimas, como el coeficiente de absorción molar y el *p*H óptimo, indicaban en el caso de nuestras ADH que éstas eran similares en todas las especies estudiadas. Además, algunas especies, así *D. simulans* y *D. hydei*, mostraban dos máximos de *p*H, uno próximo al *p*H fisiológico y el otro muy superior. (Se entiende por *p*H fisiológico aquel cuyo valor se acerca a 7,0.) No se conoce el significado biológico de unos *p*H óptimos tan básicos y tan alejados del *p*H celular pero, dado que se ignora la localización citológica de esta enzima, cualquier hipótesis resultaría arriesgada.

Las enzimas se caracterizan también a través de su punto isoeléctrico o valor de *p*H al cual la proteína no presenta carga eléctrica neta. El punto isoeléctrico depende principalmente de los grupos amino y carboxilo de la molécula. Un punto isoeléctrico elevado sugiere la presencia de muchos grupos amino y un punto isoeléctrico bajo manifiesta, por el contrario, un exceso de grupos carboxilo. La técnica de electroenfoque, que nos permite medir el punto isoeléctrico de las enzimas, consiste en establecer un gradiente de *p*H a lo largo de un soporte de poliacrilamida, aplicar la muestra en el centro del gel y someterla a un campo eléctrico. La enzima migra hasta que encuentra el *p*H de su punto isoeléctrico, adquiere entonces una carga neutra y queda inmóvil en el gel. El punto isoeléctrico es básico en todas las especies estudiadas. Y, a diferencia de lo que ocurre con la mayoría de enzimas, éstas son activas a valores de *p*H iguales a sus puntos isoeléctricos y, por tanto, detectables por incubación con el substrato adecuado en los geles de electroenfoque. Las pruebas realizadas para determinar el tamaño de la molécula, filtración en gel y electroforesis en condiciones desnaturalizantes en geles de poliacrilamida, indican que éste coincide en todas las especies. Los resultados obtenidos demuestran que la enzima activa es un dímero formado por dos subunidades idénticas de peso molecular de unos 27.000 dalton.

La técnica de electroforesis puso de manifiesto en los años sesenta la gran variabilidad genética, en el dominio de los polimorfismos enzimáticos, existente en poblaciones naturales de algunas especies de *Drosophila*. Se inició así el estudio del sistema ADH con la descripción de un sistema polimórfico en el locus *Adh* en *D. melanogaster*. De las tres variantes isoalélicas presentes en este locus Adh^S, Adh^F y Adh^{UF}, al menos dos de ellas, Adh^S y Adh^F, coexistían en las poblaciones naturales de esta especie. El mantenimiento de este polimorfismo se analizó a la luz de las hipótesis neutralistas y seleccionistas. Según los neutralistas, la variabilidad deberíase a la existencia de mutaciones que no inciden en la supervivencia y la reproducción del organismo, mientras que los seleccionistas explican el mantenimiento de las variantes genéticas por las ventajas que confieren a los individuos que las portan. En nuestro caso, diversas condiciones ambientales favorecerían las distintas aloenzimas de la ADH. La correlación existente entre la actividad alcoholdeshidrogenasa y la tolerancia al etanol e isopropanol convirtió al sistema ADH de *Drosophila* en un candidato interesante con quien demostrar la veracidad de las hipótesis seleccionistas.

Las propiedades cinéticas de las distintas variantes genéticas de la enzima ADH de *D. melanogaster* han sido determinadas por varios autores. La enzima ADH-S es menos activa frente los substratos más comunes, etanol e isopropanol, y más termoestable que la enzima ADH-F, que presenta mayor actividad enzimática frente a estos mismos substratos. La segunda es también más termolábil.

A partir de análisis electroforéticos de las variantes enzimáticas de la ADH se ha observado la existencia de una clina latitudinal, en la que el alelo Adh^S predomina en las zonas más cálidas y el Adh^F en las más frías. En nuestro caso, las especies elegidas, excepto *D. melanogaster*, resultaron ser monomórficas para la enzima ADH, y no podía realizarse este tipo de análisis. Sí cabía determinar la actividad enzimática con respecto a distintos substratos y relacionarla con el hábitat natural donde medran dichas especies. Disponíamos, además, de gran cantidad de datos, obtenidos principalmente por Antonio Prevosti y María Monclús, de nuestro departamento, que ilustraban la frecuencia de las distintas especies en bodegas y cavas en comparación con los hábitats naturales y domésticos.

El análisis de las constantes cinéticas reveló varias notas de interés. Todas las ADH estudiadas presentaban una afinidad elevada para el NAD^+ y los valores de *Km* obtenidos eran muy parecidos. (La *Km* es una constante peculiar de cada enzima, que define la concentración de substrato necesaria para alcanzar el 50 por ciento de la velocidad máxima de reacción; desde el punto de vista biológico se cree que está relacionada con la concentración intracelular de substrato. Además, junto con otros parámetros determinables *in vitro* –velocidad máxima y concentración real de substrato–, nos sirve para verificar su actividad "fisiológica", es decir, si cataliza *in vivo* la transformación de dicho substrato.) Cuando se utilizaron distintos alcoholes como substrato, manteniendo constante la concentración de NAD^+, se observó otra propiedad común para todas las especies analizadas: mayor afinidad y velocidad máxima de oxidación ante alcoholes secundarios que alcoholes primarios. No se conoce, por ahora, el significado biológico de estas características, ya que los alcoholes secundarios no son componentes frecuentes en los medios naturales donde se capturan estas especies.

A la luz de estos resultados, ¿cómo explicar los diversos grados de tolerancia al alcohol de las distintas especies analizadas? Comprobamos que existía una correlación positiva entre tolerancia al alcohol y actividad alcoholdeshidrogenásica, lo que indicaba la relevancia de esta enzima en el metabolismo del alcohol. Así, la especie más frecuentemente capturada en bodegas, *D. lebanonensis*, presentaba la máxima actividad enzimática y *D. immigrans*, especie prácticamente ausente en medios ricos en alcohol, ofrecía la mínima. Entre ellas se encontraban las otras especies en el orden esperado según dicha correlación actividad-tolerancia.

ESPECIE	COEFICIENTE DE ABSORCION MOLAR	PH OPTIMO	PUNTO ISO-ELEC-TRICO	PESO MOLECULAR (SUBUNIDAD)	EXTREMO C-TERMINAL	ACTIVIDAD ESPECIFICA (UNIDADES POR MILIGRAMO)	PROTEINA (MILIGRAMOS)
D. MELANOGASTER (*Adh^S*)	$9{,}52 \times 10^4$	8,0	7,56 7,03 6,67	27.800	-Asp-Ser-Gly-Ile[B]	22,7	13,3
D. SIMULANS	$10{,}80 \times 10^4$	7,5 9,9	8,28 8,01 7,63	27.800	-Asp-Ser-Gly-Ile	23,4	8,1
D. VIRILIS	$7{,}94 \times 10^4$	8,0 9,5	8,01 7,56 7,43	27.800	-Asp-Ser-His-Ile	17,6	17,3
D. FUNEBRIS	$21{,}94 \times 10^4$	8,6	8,60 7,99	27.600		102,4	0,5
D. IMMIGRANS	$13{,}54 \times 10^4$	8,7	9,02 8,57	27.600		21,9	3,6
D. LEBANONENSIS[A]	$13{,}30 \times 10^4$		7,95 7,50 7,15	27.800	-Asp-Ser-His-Ile	13,3	34,4
D. HYDEI	$10{,}84 \times 10^4$	7,9 9,4	8,83 8,41	27.200	-Ala (Ser / Gly) -Ile	18,0	7,8

3. PROPIEDADES BIOQUIMICAS DE LA ADH en las especies de *Drosophila* estudiadas; presentan claras semejanzas. El coeficiente de absorción molar es una constante peculiar de cada enzima y depende de la composición de aminoácidos y de la estructura terciaria de la molécula. El *p*H óptimo de las enzimas es siempre superior a 7, lo que refleja su preferencia por medios básicos. Algunas especies presentan dos *p*H óptimos; se desconoce la significación biológica de este fenómeno. La ADH activa, constituida por dos subunidades idénticas del monómero, puede presentar distintas formas moleculares que se caracterizan cada una de ellas por su punto isoeléctrico. Las bandas que se visualizan sobre el gel de electroenfoque muestran distinta intensidad en función de la proporción relativa de las diferentes formas moleculares de la enzima. En nuestro caso, la banda más intensa que corresponde a la forma predominante aparece subrayada. El peso molecular del monómero de la ADH se ha obtenido mediante electroforesis en placa de poliacrilamida con SDS; no varía en las especies estudiadas. Para determinar los residuos de aminoácidos que componen el extremo *C*-terminal de la ADH se realizó una digestión controlada con carboxipeptidasa A. El aminoácido *C*-terminal es en todos los casos analizados isoleucina, el más frecuente en tercera posición es serina y en cuarta ácido aspártico. La actividad específica indica la capacidad de la enzima para oxidar el alcohol que se utiliza como substrato, en nuestro caso, isopropanol; tómase como referencia una cantidad constante de enzima pura. Todas las actividades específicas son del mismo orden de magnitud, excepto *D. funebris*, que es unas cinco veces superior. La cantidad de proteína recuperada después de procesos de purificación equivalentes y partiendo siempre de 100 gramos de peso húmedo inicial varía mucho según la especie. Existe una buena correlación entre cantidad de enzima ADH y tolerancia a los alcoholes. Así por ejemplo, *D. lebanonensis*, la especie que mejor tolera el alcohol, presenta 2,6 veces más ADH que *D. melanogaster*, y 10 veces más que *D. immigrans*. Por su parte, *D. funebris* constituye una excepción ya que se captura con bastante frecuencia en medios ricos en alcohol; la poca cantidad de alcoholdeshidrogenasa que presenta se compensa con la elevada actividad específica de la misma. Los datos son de las autoras, salvo los señalados por (*A*), que son de las autoras y de Olof Winberg y John McKinley-Mckee, del Instituto de Bioquímica de la Universidad de Oslo, y los señalos por (*B*), que son de David Thatcher, cuando trabajaba en la Universidad de Edimburgo.

Después de realizar varias purificaciones partiendo siempre del mismo peso de individuos adultos y siguiendo idéntico protocolo experimental, se observó que las cantidades de enzima pura obtenidas variaban según la especie. Las variaciones en cantidad de enzima eran muy superiores a las encontradas en actividad específica. En todas las especies existía una buena correlación entre cantidad de enzima y tolerancia al alcohol, con la única excepción de *D. funebris* es difícil valorar de modo preciso la cantidad de enzima ADH presente en individuos de distintas especies, si bien los resultados mencionados, obtenidos mediante el proceso de purificación, constituyen una buena indicación de la causa determinante de la tolerancia al alcohol. El caso de *D. funebris* es una excepción interesante, pues aunque sintetiza una cantidad de enzima muy reducida, ésta se compensa con una elevada actividad específica, del orden de 5 veces superior a la de las otras especies. De modo que, si bien la adaptación a medios más ricos en alcohol se explica por un incremento en la cantidad de enzima, en *D. funebris* se alcanza el mismo efecto gracias al incremento substancial de actividad específica.

Determinamos la composición de aminoácidos de la enzima ADH a partir de muestras de enzima pura hidrolizadas en medio ácido a 105 grados Celsius. Los resultados revelaron un elevado grado de homología entre las distintas especies. En particular, todas las ADH poseen dos residuos de cisteína por subunidad. El índice o parámetro de Bhattacharyya nos sirvió para establecer las relaciones filogenéticas entre las especies estudiadas basadas en la homología de la ADH. Dicho parámetro se simboliza formalmente así: $d_{\alpha\beta} = 2 \text{ arc } cos\ (\Sigma \sqrt{p_i q_i})$, donde $p_1 \ldots p_n$ son las frecuencias relativas de los aminoácidos de la ADH en la especie α y $q_1 \ldots q_n$ las frecuencias relativas de los aminoácidos de la especie β. (El sumatorio Σ abarca desde $i=1$ hasta $i=n$.) Utilizamos esta distancia porque, desde el punto de vista geométrico, designa la distancia geodésica natural, ya que la curvatura del espacio no es constante.

Según el dendrograma obtenido, las siete especies aparecían agrupadas en tres ramas: *D. melanogaster* y *D. simulans; D. virilis* y *D. lebanonensis; D. immigrans, D. funebris* y *D. hydei*. Esta distribución estaba de acuerdo con las relaciones filogenéticas establecidas en *Drosophila* por Patterson y Stone en 1952 después de analizar un elevado número de caracteres morfológicos y del estudio de los cromosomas. *D. melanogaster* y *D. simulans* son especies gemelas que pertenecen al subgénero *Sophophora*; dentro del grupo *melanogaster* son, sin duda, las especies más próximas. *D. hydei, D. funebris* y *D. immigrans* pertenecen al subgénero *Drosophila* y están consideradas como las especies ancestrales de sus respectivos grupos en razón de la distribución geográfica que presentan. *D. virilis*, si bien pertenece al mismo subgénero, no es cosmopolita; su pariente más próximo en el mismo subgénero es *D. pinicola*, especie muy próxima al grupo *obscura*. Así pues *D. virilis* guardaría una relación más afín con las especies del subgénero *Sophophora* que *D. hydei*, *D. funebris* o *D. immigrans*.

Nuestros resultados concordaban con esta relación. El origen de *D. lebanonensis* está aún sujeto a controversia. Esta especie, del grupo *victoria*, podría, según algunos autores, proceder del grupo *repleta* o *virilis*, al que pertenecen *D. hydei* y *D. virilis*, respectivamente. Otros estudios sitúan a

este grupo próximo al origen de las radiaciones del género y por tanto bastante alejado de las especies mencionadas. Las distancias filogenéticas deducidas a partir de la composición de aminoácidos de la ADH apoyarían la primera hipótesis.

Decidimos iniciar el estudio de la estructura de las ADH mediante la construcción de mapas peptídicos: separación bidimensional sobre papel de los péptidos obtenidos después de una digestión hidrolítica. Se partía de enzima pura, que se desnaturalizaba totalmente para que fuese accesible a la enzima hidrolítica; realizábase luego una separación bidimensional de los péptidos obtenidos: la primera por electroforesis y la segunda por cromatografía ascendente a 90 grados de la primera. Los péptidos se visualizaban por tinción con fluorescamina o con otros colorantes específicos para algunos aminoácidos. La comparación de los mapas analíticos no revelaba diferencias importantes entre las especies elegidas. Ello se debía, principalmente, a que la separación de los péptidos no era total; suele darse, por contra, un solapamiento, que enmascara así las posibles diferencias. Por otra parte, sólo algunas substituciones de aminoácidos alterarían la movilidad electroforética o la hidrofobicidad del péptido; además, estos efectos serían difícilmente detectables en los casos en que su tamaño fuera elevado, es decir, superior a seis o siete aminoácidos de longitud.

En el caso de *D. hydei* decidimos elaborar un mapa peptídico preparativo de la ADH, eluir los péptidos obtenidos y proceder a una purificación posterior mediante cromatografía líquida de alta resolución. Esto equivalía a una filtración a alta presión a través de bolas de sílice de cinco micrometros de diámetro, cubiertas por una capa hidrofóbica formada por moléculas de 18 átomos de carbono.

De cada péptido inicial pudimos aislar el componente mayoritario, para analizar después los aminoácidos y determinar el residuo *N*-terminal, requisitos indispensables para su posterior identificación. Comparamos estos péptidos con los de *D. melanogaster* Adh^S, cuya secuencia primaria había sido establecida por David Thatcher, en 1980, en la Universidad de Edimburgo. De los 255 aminoácidos que constituyen el monómero de la ADH de *D. melanogaster* Adh^S sólo pudimos identificar 160 correspondientes a *D. hydei*. Teniendo en cuenta que este número de residuos equivale al 64 por ciento del total de aminoácidos que constituyen el monómero, encontramos un 21,8 por ciento de substituciones distribuidas uniformemente en la mitad *N*-terminal y en la mitad *C*-terminal de la cadena peptídica. Este 21,8 por ciento equivaldría, por un lado, a una estima mínima de las substituciones de aminoácidos de la ADH y, por otro, no todas

ACETIL-Ser-Phe-Thr-Leu-Thr-Asn-Lys-Asn-Val-Ile-Phe-Val-Ala-Gly-Leu-Gly-Gly-Ile-Gly-Leu-Asp-Thr-Ser-Lys-Glu-Leu-Leu-Lys-Arg-Asp-Leu-Lys-Asn-

Asp,Asp,Leu,Lys | Asp

Leu-Val-Ile-Leu-Asp-Arg-Ile-Glu-Asn-Pro-Ala-Ala-Ile-Ala-Glu-Leu-Lys-Ala-Ile-Asn-Pro-Lys-Val-Thr-Val-Thr-Phe-Tyr-Pro-Tyr-Asp-Val-Thr-Val-Pro-Ile-Ala-

Leu,Gly,Pro,Leu,Asp,Arg | Ile,Glu,Asp,Pro,Ala,Ala,Ile,Ala,Glu,Leu,Lys | Glu,Ile,Glu,Val,Lys

Glu-Thr-Thr-Lys-Leu-Leu-Lys-Thr-Ile-Phe-Ala-Gln-Leu-Lys-Thr-Val-Asp-Val-Leu-Ile-Asn-Gly-Ala-Gly-Ile-Leu-Asp-Asp-His-Gln-Ile-Glu-Arg-Thr-Ile-Ala-

Leu,Leu,Lys | Thr,Val,Asp,Ala,Leu,Ile,Asp,Gly,Ala,Gly,Ile,Leu | Asp,Tyr,Ser,Glu,Pro,Glu,Arg

Val-Asn-Tyr-Thr-Gly-Leu-Val-Asn-Thr-Thr-Thr-Ala-Ile-Leu-Asp-Phe-Thr-Asp-Lys-Arg-Lys-Gly-Gly-Pro-Gly-Gly-Ile-Ile-Cys-Asn-Ile-Gly-Ser-Val-Thr-Gly-

Gly,Gly,Pro,Gly | Gly,Ile,Ile,Cys,Asp,Asp,Gly,Asp | Val,Thr,Gly,

Phe-Asn-Ala-Ile-Tyr-Gln-Val-Pro-Val-Tyr-Ser-Gly-Thr-Lys-Ala-Ala-Val-Val-Asn-Phe-Thr-Ser-Ser-Leu-Ala-Lys-Leu-Ala-Pro-Ile-Thr-Gly-Val-Thr-Ala-Tyr-

Phe,Asp,Ala,Ile,Tyr,Glu,Val,Pro,Leu,Asp,Ser,Gly,Thr,Lys | Ala,Ala,Ala,Leu,Asp,Phe,Thr,Ser,Ser,Leu,Ala,Lys | Leu,Ala,Pro,Ile,Thr,Gly,Val,Thr,Ala,Tyr

Thr-Val-Asn-Pro-Gly-Ile-Thr-Arg-Thr-Thr-Leu-Val-His-Lys-Phe-Asn-Ser-Trp-Leu-Asp-Val-Glu-Pro-Gln-Val-Ala-Glu-Lys-Leu-Leu-Ala-His-Pro-Thr-Gln-Pro

Ser,Ile,Asp,Pro,Gly,Ile,Thr,Lys | Thr,Thr,Leu,Val,His,Lys | Phe,Asp,Ile,(Trp),Leu,Asp,Val,Gly,Asp,Lys | Leu,Ala,Glu,Lys | Leu,Val,Ala,His,Pro,Ile,Glu,Pro,

-Ser-Leu-Ala-Cys-Ala-Glu-Asn-Phe-Val-Lys-Ala-Ile-Glu-Leu-Asn-Gln-Asn-Gly-Ala-Ile-Trp-Lys-Leu-Asp-Leu-Gly-Thr-Leu-Glu-Ala-Ile-Gln-Trp-Thr-Lys-His

Ser,Val,Ala,Cys,Ala,Lys | Asp,Ser,Val,Glu,Ala,Ile,Glu,Leu,Glu,Lys | Ala,Ala,His,

-Trp-Asp-Ser-Gly-Ile COOH.

(Trp),Asp,Ser,Gly,Ille COOH.

ADH DE *DROSOPHILA MELANOGASTER*

ADH DE *DROSOPHILA HYDEI*

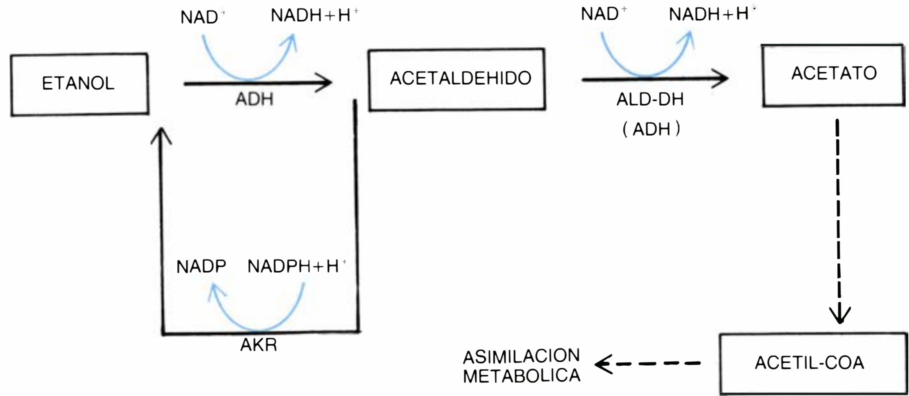
4. SECUENCIA DE AMINOACIDOS de una de las dos subunidades idénticas que dan lugar al dímero activo de la enzima alcoholdeshidrogenasa de *Drosophila melanogaster* (variante Adh^S), publicada en 1980 por David Thatcher, de la Universidad de Edimburgo. El grupo amino que quedaría libre en uno de los extremos de la molécula (extremo *N*-terminal) se encuentra enlazado a un radical acetilo. En el otro extremo (*C*-terminal), una isoleucina presenta un grupo carboxilo libre. En un segundo nivel aparece la composición de aminoácidos de los péptidos de la ADH de *D. hydei*, purificados después de una digestión tríptica. La tripsina es una enzima que hidroliza los enlaces peptídicos, en los cuales el grupo carbonilo ha sido aportado por una lisina o una arginina. Además de su composición, se determinó, para cada péptido, el aminoácido *N*-terminal, que aparece destacado. Existe una ambigüedad en los residuos de ácido aspártico y ácido glutámico, ya que alguno de ellos debe corresponder a asparagina o glutamina, respectivamente. Los distintos aminoácidos se hallan separados por comas, pues no se realizó la secuenciación de los péptidos, sino que, a partir de su composición, se identificó cada péptido de la alcoholdeshidrogenasa de *D. hydei* con el de la alcoholdeshidrogenasa de *D. melanogaster* (Adh^S) que mostraba mayor coincidencia. Los residuos de aminoácidos se ordenaron de acuerdo con ese mismo criterio. De este modo, se identificaron 160 aminoácidos, lo que constituye un 64 por ciento del total del monómero. Encontramos un 21,8 por ciento de substituciones distribuidas en la mitad *N*-terminal y *C*-terminal de modo uniforme. Obsérvese cómo la mayor parte de los péptidos recuperados después del proceso de purificación corresponden a la mitad *C*-terminal de la molécula. Dado que los criterios de la purificación se basan sobre todo en la solubilidad de los péptidos, este hecho indicaría un carácter hidrofílico de esta porción de la molécula, frente a una marcada hidrofobicidad de la mitad *N*-terminal.

las substituciones serían detectables con las técnicas utilizadas.

Al considerar las propiedades bioquímicas y especialmente las cinéticas de la ADH de *D. hydei* se observan semejanzas claras con *D. simulans*, más que con ninguna de las demás especies. Bodmer y Ashburner, de la Universidad de Cambridge, secuenciaron en 1984 el gen *Adh* de un conjunto de especies próximas del grupo *melanogaster*, y entre ellas *D. simulans*. A partir de la secuencia de nucleótidos se pueden predecir las substituciones de aminoácidos en la proteína y comparar así *D. melanogaster* Adh^S y *D. simulans*. Las substituciones descritas son sólo dos, una serina que remplaza a una alanina en posición 2 y otra glutamina que hace lo propio con una lisina en posición 83. Estas dos substituciones inciden notablemente en las propiedades catalíticas de la enzima ADH y acaban por producir un efecto semejante en el conjunto de substituciones entre *D. melanogaster* Adh^S y *D. hydei*. Así pues, parece que gran parte de las substituciones que han tenido lugar en este gen de *D. hydei* no se reflejan en las propiedades catalíticas de la molécula, ni probablemente alteran su función fisiológica.

Otra de las cuestiones que nos planteamos al iniciar este proyecto de investigación se refería a la función de la enzima ADH en relación con otras enzimas implicadas en el metabolismo del alcohol. La ADH cataliza en *Drosophila*, al igual que en otras especies animales, la oxidación de los alcoholes.

ACTIVIDAD ENZIMATICA (UNIDADES POR MILILITRO)
ADH (CONTROL)
AKR
ADH
AKR (CONTROL)
ETANOL (CONTROL)
ETANOL
ETANOL INTERNO (× 100)
ACETALDEHIDO
CONCENTRACION EN MILIGRAMOS POR 50 MILIGRAMOS DE INDIVIDUOS ADULTOS
TIEMPO EN HORAS

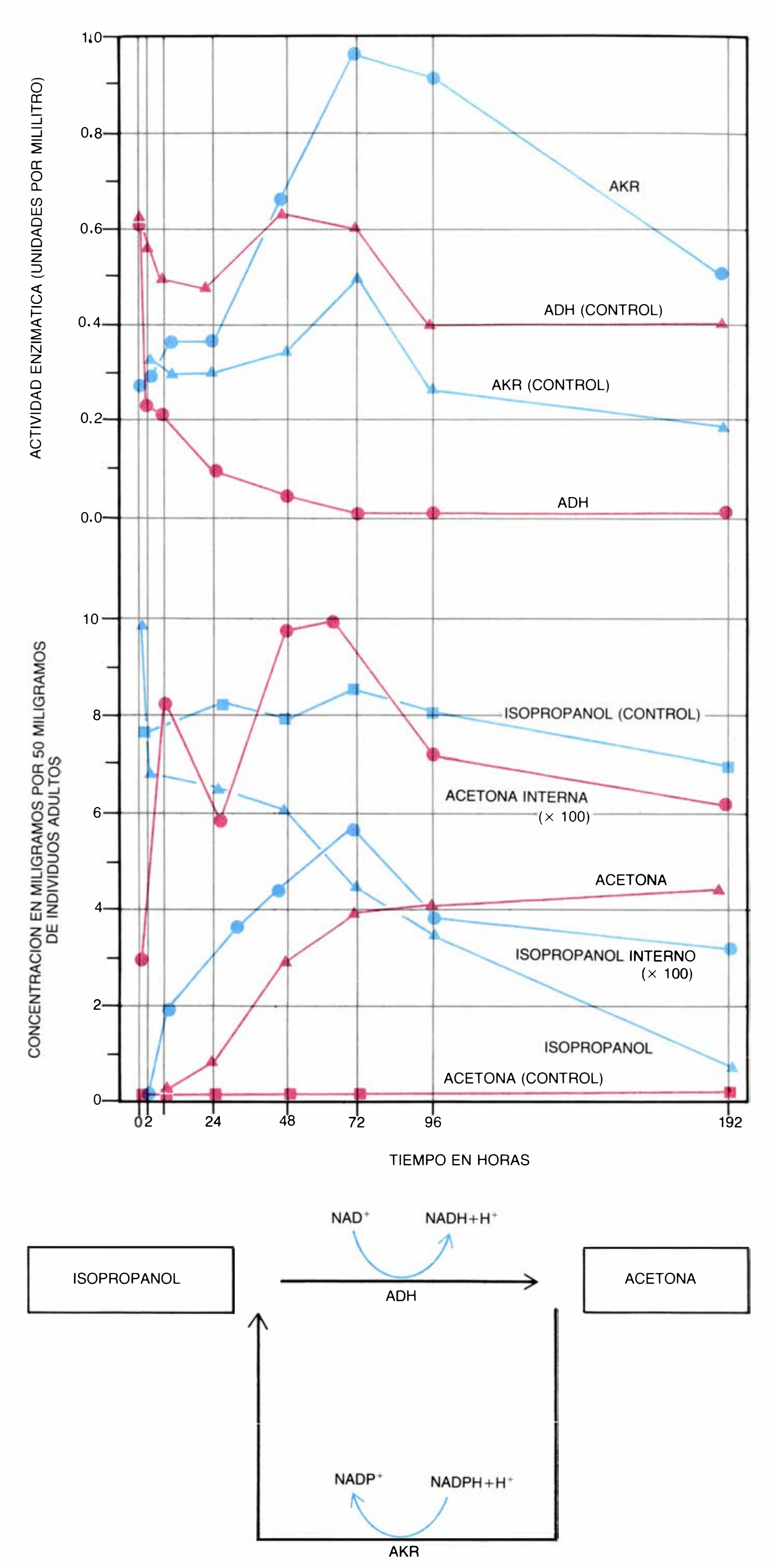


5. LA ENZIMA ADH DE *DROSOPHILA* interviene en el metabolismo del etanol, ya que cataliza su conversión en acetaldehído, que rápidamente se transforma en acetato y luego en acetil-CoA. En la transformación del acetaldehído intervienen otras enzimas: el paso a acetato está controlado por la aldehidodeshidrogenasa (ALD-DH) y también, en algunas especies de *Drosophila*, por la alcoholdeshidrogenasa (ADH). En *D. melanogaster* y *D. hydei*, el acetaldehído puede transformarse otra vez en etanol mediante el concurso de la aldo-cetorreductasa (AKR). Cuando se transfieren individuos adultos de 3-4 días de *D. hydei* a un frasco de cultivo que contiene en el fondo una celulosa empapada con una solución de etanol al 1 por ciento en sacarosa al 3 por ciento, se detecta un incremento de la actividad ADH y AKR. Los valores máximos se alcanzan a las 48 horas de tratamiento. Transcurrido ese intervalo, ambas actividades inician un descenso progresivo. Por cromatografía de gases puede medirse la concentración interna y externa de etanol y acetaldehído a lo largo del experimento. Se observa que el etanol desciende progresivamente en el medio de cultivo a medida que lo incorporan los individuos. La concentración de etanol interno presenta un máximo hacia las 8 horas de tratamiento, fluctúa ligeramente durante un tiempo y se inicia un descenso a los cuatro días, cuando ya se ha agotado el etanol externo.

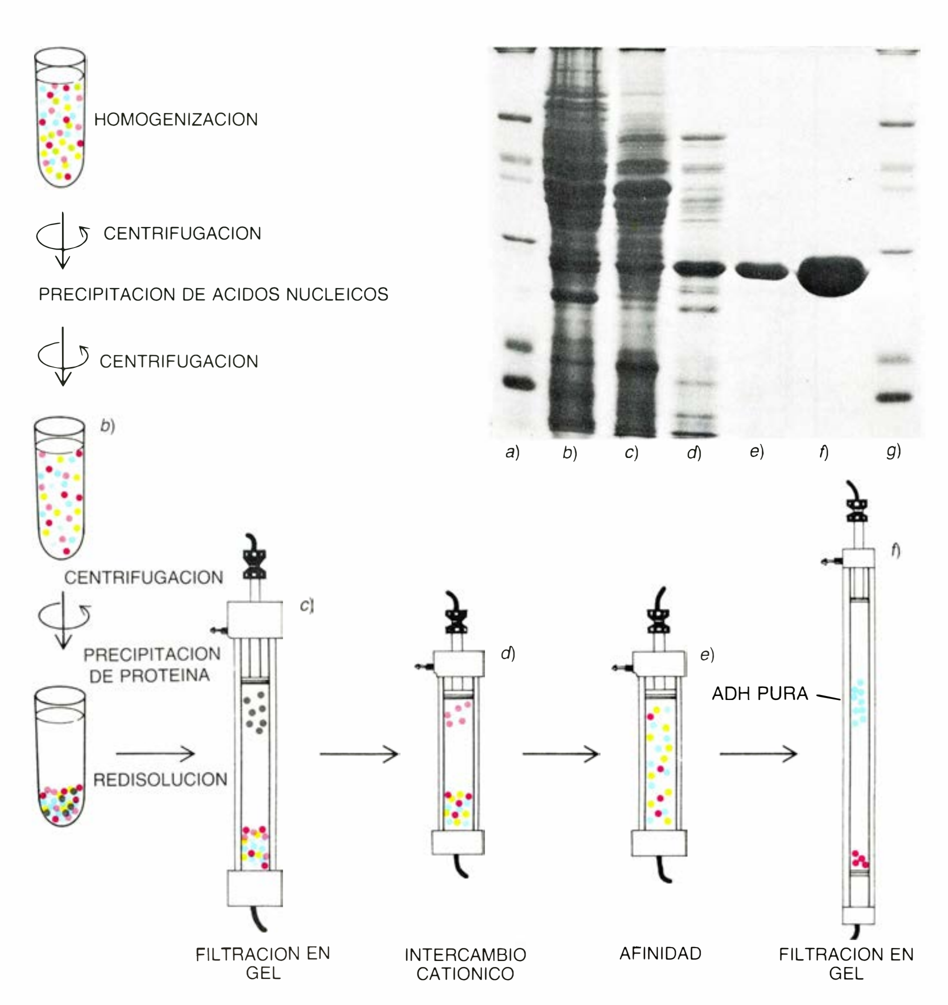
Para ello, usa NAD$^+$ como cofactor. Llámase unidad de actividad enzimática la cantidad de enzima necesaria para reducir 1 micromol de NAD$^+$ por minuto a 25 grados Celsius. Esta reacción puede cuantificarse espectrofotométricamente midiendo el incremento de absorbancia que se produce a 340 nanómetros debido a la reducción del NAD$^+$. También puede cuantificarse la reacción inversa, es decir, la oxidación del NADH.

El efecto de la ingestión de alcoholes, o de sus productos de oxidación, en *Drosophila* varía según las dosis utilizadas o la naturaleza del producto en cuestión. A dosis bajas de etanol (1-2 por ciento), se prolonga la vida media de los individuos en ausencia de otros nutrientes, mientras que, a dosis elevadas, aparecen los efectos tóxicos primarios y secundarios que causan la muerte. El isopropanol a concentraciones bajas (0,5-1 por ciento) retrasa el ciclo biológico del organismo y, a dosis más elevadas, produce la muerte.

Estos datos obtenidos en *D. melanogaster* por el grupo de Jean David, del laboratorio de biología y genética evolutiva del Centro Nacional francés de Investigaciones Científicas en Gif-sur-Yvette, y de Wilke Van Delden, del departamento de genética de la Universidad holandesa de Groningen, a finales de los años setenta, junto con los de nuestro grupo de trabajo con *D. immigrans, D. funebris, D. hydei* y *D. melanogaster*, ponen de manifiesto la doble función de la enzima ADH. A dosis limitadas de etanol, la ADH suministra productos que se incorporan



6. EFECTO DE ISOPROPANOL en *Drosophila*. Difiere notablemente del ejercido por el etanol. Los individuos adultos de *D. hydei* transferidos a un medio que contiene un 0,5 por ciento de isopropanol en sacarosa al 3 por ciento presentan un fuerte incremento de actividad AKR y un descenso rápido de la actividad ADH. La enzima ADH cataliza la conversión de isopropanol en acetona, compuesto que tiende a acumularse como tal y no sufre una metabolización posterior en el interior del organismo. La enzima AKR utiliza, en cambio, la acetona como substrato, y logra reducirla a isopropanol. A los tres días de tratamiento se obtiene una doble respuesta: máxima actividad AKR y mínima de la ADH. La AKR máxima coincide a su vez con la máxima concentración interna de acetona y la mínima de ADH con la máxima de isopropanol. La evolución de ambas actividades enzimáticas ofrece al individuo la capacidad de una respuesta rápida a concentraciones internas de isopropanol y acetona, constituyendo un mecanismo adaptativo que le permite sobrevivir en un medio poco favorable. La acetona acaba excretándose al medio externo. A los ocho días de tratamiento, el isopropanol externo ha desaparecido; por el contrario, la concentración externa de acetona es máxima. Las concentraciones internas de isopropanol y acetona se mantienen bajas, alcanzándose niveles superiores de acetona que de isopropanol.



7. PURIFICACION DE LA ENZIMA ALCOHOLDESHIDROGENASA (ADH). Se realiza a partir de 100 gramos de *Drosophila* que se homogenizan en tampón-fosfato a un *p*H adecuado para que la carga neta de la enzima sea positiva. Este paso tiene por objeto liberar la enzima del medio intracelular y ponerla en solución. Seguidamente, se eliminan los ácidos nucleicos que pueda haber en la preparación, precipitándolos con sulfato de protamina y posterior centrifugación. La proteína disuelta en el sobrenadante (*b*) se precipita con sulfato amónico para reducir el volumen. El sedimento obtenido después de centrifugar se redisuelve en tampón-fosfato; los iones se separan de la proteína por filtración en gel. La solución de proteína resultante (*c*) se aplica a una columna de intercambio catiónico débil, donde queda retenida; su elución se realiza con un gradiente de cloruro sódico, obteniéndose la preparación (*d*). Esta se aplica a una columna de afinidad que posee como ligando un análogo del NAD^+. La alcoholdeshidrogenasa retenida en esta columna se eluye con NAD^+; se logra una preparación casi pura (*e*). El NAD^+ y alguna proteína contaminante de mayor peso molecular que la ADH se eliminan finalmente por filtración en gel. La preparación que se obtiene contiene sólo ADH, como puede apreciarse en el paso *f*. La fotografía corresponde a un gel de acrilamida con SDS (siglas de dodecil sulfato de sodio), producto que desnaturaliza las proteínas y les confiere una carga neta negativa. Las columnas (*a*) y (*g*) presentan la separación de una mezcla de proteínas convencionales.

en el metabolismo energético, mientras que a dosis más elevadas o frente a alcoholes secundarios su función es principalmente detoxificante, evitando que se cree un exceso del producto secundario, que en general, tal como sucede en el caso del acetaldehído o la acetona resulta tan tóxico, sino más, que el propio alcohol. Esta segunda propiedad de la enzima se realiza por inhibición de su propia actividad enzimática y va acompañada de un intercambio en la proporción de las distintas formas moleculares asociadas a la ADH. Este efecto se detecta por electroforesis; el descenso de actividad enzimática se determina por métodos espectrofotométricos, cuantificando la reducción del NAD^+.

Decidimos realizar una serie de experimentos para analizar el efecto de los alcoholes. El efecto metabólico del etanol a bajas concentraciones variaba según la actividad alcoholdeshidrogenásica de la especie. Se favorecía el desarrollo larvario en un medio al que se había suministrado etanol en el caso de *D. melanogaster* y *D. funebris*; reducíase en el caso de *D. immigrans*. La menguada viabilidad larvaria en *D. immigrans* era debida, seguramente, a la baja actividad de la ADH que determinaba una acumulación del etanol, alcanzándose niveles tóxicos en el interior del organismo. El efecto del isopropanol resultaba nocivo para el desarrollo larvario en las especies citadas. Además, podía estimarse la cantidad de isopropanol y acetona en el medio de cultivo por cromatografía de gases a distintos tiempos de desarrollo.

Cuando realizamos estas medidas, observamos que la concentración de acetona, en individuos que se habían desarrollado en un medio dotado de isopropanol al 0,5 por ciento, se incrementaba hasta alcanzar un valor máximo, a partir del cual volvía a transformarse en isopropanol. Este proceso adaptativo ocurría varias veces a lo largo del desarrollo larvario de *D. melanogaster* Adh^S, *D. funebris* y *D. immigrans*. Nos preguntamos cuál era la actividad enzimática responsable de la reducción de la acetona en isopropanol. Al cuantificar la actividad inversa de la ADH se demostró que ésta no podía explicar la cantidad de acetona resultante, debido, entre otros factores, a la propia inhibición de la actividad alcoholdeshidrogenásica que se producía en los individuos que crecían en presencia de isopropanol y que afectaba tanto a la actividad directa como a la inversa de la ADH.

¿Cuál era la actividad enzimática responsable de la reducción de los productos de oxidación de los alcoholes? Logramos caracterizar en *Drosophila* una actividad aldo-ceto reductásica (AKR), detectada con anterioridad en varias especies, ciertos mamíferos incluidos.

Dicha actividad enzimática requería la presencia del cofactor NADPH, y presentaba, en orden decreciente de actividad, la siguiente especificidad de substratos: acetaldehído, p-nitrobenzaldehído, D-gliceraldehído, o-nitrobenzaldehído, L-gliceraldehído y acetona. Podía cuantificarse la actividad enzimática en el espectrofótometro midiendo la oxidación del cofactor; dejaba verse, asimismo, sobre geles de almidón. La utilización del NADPH como cofactor, los substratos mencionados y la movilidad electroforética distinguían claramente la actividad AKR de la actividad alcoholdeshidrogenasa.

Sin embargo, para descartar toda sospecha de contribución de la enzima ADH decidimos recurrir a mutantes genéticos ADH-negativos; tratábase de cepas de *D. melanogaster* que carecían del gen *Adh* o bien el producto que se sintetizaba no era funcional. Demostramos que la actividad AKR era totalmente independiente de la ADH. Los niveles de actividad enzimática de las cepas mutantes para *Adh* eran iguales a las de los individuos salvajes.

Otras características bioquímicas permitían diferenciar las dos actividades enzimáticas. La AKR era sensible al fenobarbital y en cambio no se inhibía en presencia de pirazol, un potente inhibidor de la ADH de *Drosophila*. Por todo el conjunto de propiedades descritas, esta actividad enzimática era la análoga a la AKR, descubierta ya en otras especies animales.

La función biológica de esta enzima podría revestir gran interés si aporta un posible mecanismo de adaptación de la *Drosophila* a los medios ricos en alcohol. Así, en presencia de isopropanol, la acetona producida, un potente tóxico no metabolizable, en vez de acumularse en el interior del organismo podría transformarse parcialmente en isopropanol. Por otro lado, tanto la acetona como el isopropanol inhibirían la actividad alcoholdeshidrogensa evitando mayor producción de acetona. Observamos, además, que la acetona se excretaba como tal al medio de cultivo; y quizás este proceso no se desarrolle con la suficiente rapidez y la eficacia necesaria para producir, por sí solo, niveles tolerables de concentración intracelular. Así, en el proceso de adaptación al isopropanol operarían simultáneamente los dos mecanismos descritos, produciendo un mismo efecto detoxificador.

En presencia de etanol, la actividad AKR impediría la acumulación de acetaldehído, evitando que se alcanzasen las cotas peligrosas de toxicidad. En este caso, la actividad ADH se incrementa hasta pasadas las 48 horas de tratamiento, para iniciar después un descenso continuado. Este incremento de actividad ADH comporta una mayor producción de acetaldehído. Parte de este acetaldehído producido se transformaría en acetato mediante una actividad aldehidodeshidrogenasa y el resto volvería a convertirse en etanol. Así se explicaría que las concentraciones de acetaldehído intracelular y en el medio de cultivo resulten apenas detectables.

Con el fin de estudiar la evolución de la actividad ADH y aldo-ceto reductasa para verificar las hipótesis anteriores, decidimos someter individuos adultos de *D. hydei* a medios que contenían diversas concentraciones de etanol e isopropanol. Para ello, individuos jóvenes de la misma edad eran transferidos a frascos de cultivo que contenían una celulosa empapada con una solución de sacarosa y etanol o isopropanol. Se cuantificaban las actividades enzimáticas a lo largo del tratamiento. Se demostró que, en los individuos tratados con etanol, las actividades ADH y AKR seguían una evolución paralela: se incrementaban durante las primeras 48 horas de tratamiento para iniciar después un descenso progresivo. En cambio, en los individuos tratados con isopropanol se observaba un descenso rápido de actividad alcoholdeshidrogenasa y un incremento brusco de actividad AKR. Además, las actividades mínima y máxima coincidían a las 96 horas de tratamiento, que a su vez se correspondían con los niveles máximos de acetona e isopropanol. Así, pues, parecía confirmarse nuestra hipótesis inicial sobre la función biológica de la enzima AKR y su importante papel con respecto a la adaptación de algunas especies de *Drosophila* a los medios ricos en alcohol.

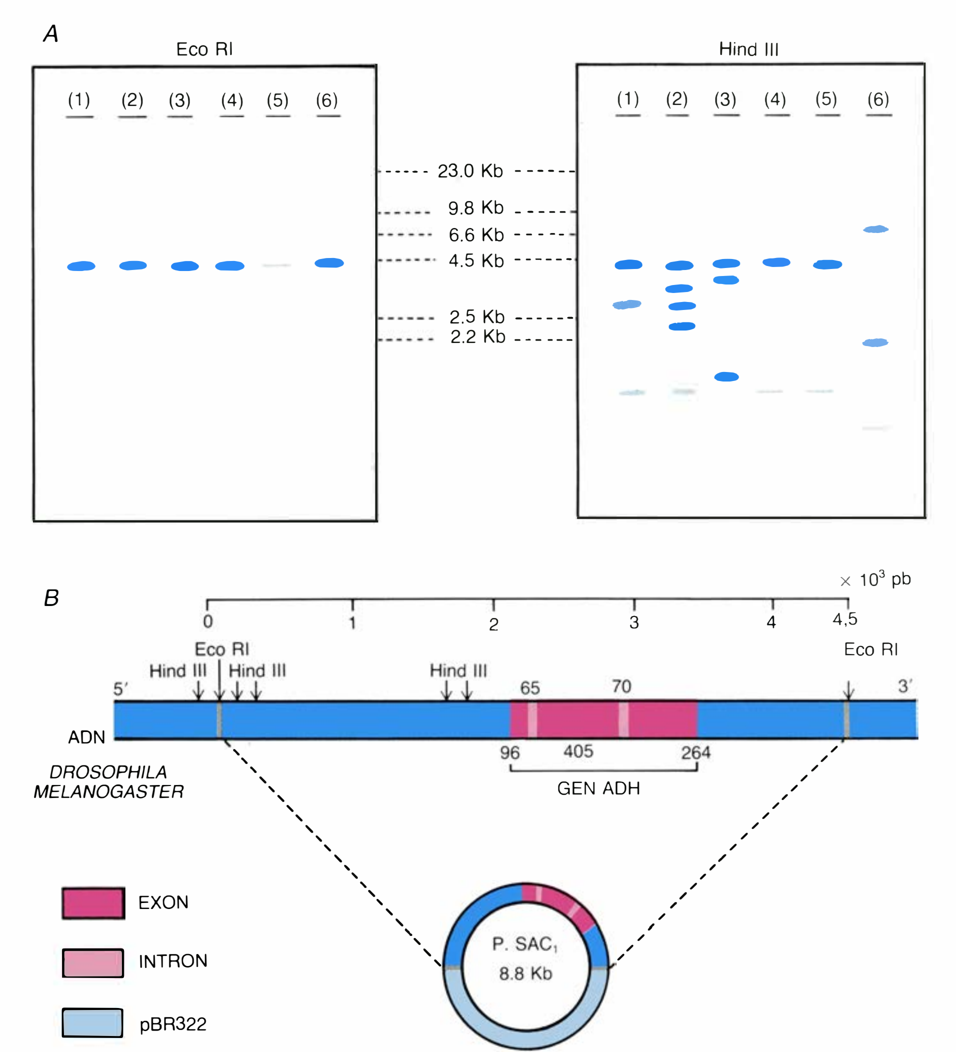
Tras someter individuos de diversas especies a un tratamiento con isopropanol y transferirlos a un medio sin alcohol se recupera la actividad ADH, hasta alcanzarse niveles de actividad enzimática próximos a los iniciales. Tal como se ha indicado anteriormente, la inhibición de actividad iba siempre acompañada de un cambio en las proporciones relativas de las formas moleculares de la ADH; en presencia de isopropanol se favorecían las formas más anódicas, que eran las más inactivas. Al recuperarse la actividad ADH volvía a aparecer el patrón de bandas normal. Estos procesos de interconversión se detectaban en electroforesis sobre geles de almidón y en electroenfoque.

No se conocían las causas moleculares de este fenómeno de recuperación de la actividad, descrito en *D. melanogaster*, *D. funebris*, *D. hydei*, *D. immigrans* y *D. buzzatii*. Había dos posibilidades: la recuperación de actividad ADH debíase a la reactivación de moléculas ya existentes o era instada por nueva síntesis de enzima. Para verificar qué hipótesis era la correcta sometimos individuos adultos de *D. hydei* a un tratamiento con isopropanol durante 24 horas. Transcurrido ese tiempo, se transfirieron a un frasco de cultivo que contenía cicloheximida. La cicloheximida es un inhibidor de la síntesis de proteínas. (El grupo de John Maynard Smith, de la Universidad de Sussex, había desentrañado en 1968 las características de esta inhibición en *Drosophila*.) En nuestro caso, en presencia de cicloheximida se inhibía totalmente la recuperación de actividad ADH, mientras que en los controles, que se mantenían en ausencia de cicloheximida, se alcanzaban los niveles de actividad esperados. Se demostraba así que la recuperación de actividad debíase a una nueva síntesis de enzima ADH.

El estudio de la evolución de la enzima ADH había puesto de manifiesto las limitaciones del análisis del producto génico: una fracción importante de las substituciones en los tripletes que determinan (codifican) los distintos aminoácidos que constituyen la proteína no se reflejaría finalmente en la cadena peptídica. Las alteraciones sufridas en los intrones, o regiones del gen que no se traducen, pasarían inadvertidas; además, analizando la enzima no tendríamos ninguna información sobre las regiones del genoma próximas al gen que intervienen en el proceso de expresión del mismo. El próximo paso obligado para profundizar en la evolución y expresión del sistema ADH consistía en su estudio genómico, mediante técnicas de ingeniería genética.

A principios de la década de los ochenta, varios laboratorios abordaron el estudio de la estructura y la expresión del gen *Adh*. En 1980, David Goldberg, del Instituto de Tecnología de California, obtuvo una molécula recombinante entre *p*BR322, un plásmido de uso ya universal en ingeniería genética, y un segmento de ADN de *D. melanogaster* que comprendía el gen que codifica para la enzima ADH. La denominó plásmido *s*AC1. El marcaje de este plásmido proporciona una herramienta preciosa para los estudios del gen *Adh*. Desde la fecha indicada, importantes grupos de investigación han centrado su interés en este tema. Cabe destacar las aportaciones de Martin Kreitman en 1983, de la Universidad de Harvard, quien ha secuenciado el gen *Adh* y regiones adyacentes de *D. melanogaster*; y de Cheeptip Benyajati en 1981, en el Instituto Nacional del Cáncer de Maryland, que ha caracterizado su producto de expresión o ARN mensajero.

Se sabe así que, al igual que en la mayoría de los genes eucariotas, los exones, o secuencias de ADN que aparecerán representadas más tarde en la secuencia proteica, se hallan interrumpidos por intrones, o secuencias que resultan eliminadas durante el procesamiento del gen. En el caso que nos ocupa, el gen *Adh* de *D. melanogaster* posee dos intrones, de 65 y 70 pares de bases, que separan exones de 96, 405 y 264 pares de bases, respectivamente.



8. COMPARACION DE LAS DIANAS DE RESTRICCION en la región de ADN que contiene el gen *Adh* (*izquierda*). Se purificó el ADN de cada una de las siguientes especies de *Drosophila: D. subobscura* (1), *D. guanche* (2), *D. melanogaster* (*Adh^S*) (3), *D. madeirensis* (4) y *D. ambigua* (5). Este se digirió a continuación con una restrictasa, o enzima que hidroliza los enlaces fosfodiéster de la cadena de nucleótidos en secuencias específicas. Hemos utilizado la enzima *Eco* RI, que reconoce y corta por la secuencia G/AATTC, y la enzima *Hind* III, que lo hace por la secuencia A/AGCTT. Los fragmentos resultantes se separan por tamaños mediante electroforesis en gel de agarosa y se transfieren a un filtro de nitrocelulosa (la llamada técnica de Southern), que se hibrida con una sonda de ADN marcada radiactivamente con ^{32}P. Se usó como sonda el plásmido *s*AC1 (6), un híbrido de *p*BR322 y un segmento de 4,5 kilobases del ADN de *D. melanogaster* (*Adh^S*) que contiene el gen *Adh*, y que se halla flanqueado por dianas *Eco* RI. Las bandas que incluyen fragmentos que presentan alguna secuencia homóloga a la de la sonda se visualizan por autorradiografía. Su tamaño puede deducirse por comparación con las bandas del ADN del fago λ digerido por *Hind* III. Se observa cómo el fragmento delimitado por las dianas *Eco* RI parece conservarse en todas las especies y coincide con el de *D. melanogaster*. No ocurre así con el patrón de bandas para la enzima *Hind* III: mientras en algunas especies –*D. madeirensis* y *D. ambigua*– habría desaparecido alguna diana respecto a *D. melanogaster*, en *D. guanche* se habría incrementado su número y en *D. subobscura* habrían cambiado su distribución. A la derecha se ilustra la estructura del segmento de ADN de *D. melanogaster* (*Adh^S*) que contiene el gen *Adh*. El segmento de 4,5 kilobases flanqueado por dianas de restricción para *Eco* RI fue clonado en *p*BR322 para dar origen al plásmido *s*AC1., que se usa de sonda para detectar la presencia de este gen. En el diagrama se observa también la localización de los tres exones del gen *Adh* de *D. melanogaster* (*Adh^S*) que darán lugar a la proteína y de los dos intrones que los separan. Se indica la distribución de las dianas para la restrictasa *Hind* III.

Se han identificado también las señales o secuencias específicas de nucleótidos que dirigen el inicio de la transcripción y las sucesivas etapas de maduración de la molécula de ARN mensajero, desde su síntesis hasta su entrada en el ribosoma. En 1983, C. Benyajati, en colaboración con Michael Ashburner, de la Universidad de Cambridge, llegó a la sorprendente conclusión de que estos segmentos no eran los mismos en el ARN mensajero de las larvas de *D. melanogaster* que en el de los adultos; el resto del gen resulta transcrito a partir de una misma secuencia.

Nuestro equipo se propuso usar el plásmido *s*AC1 para sus estudios, basándose en la homología que debía existir, cuando menos en los exones, entre el gen *Adh* de *D. melanogaster* y el de otras especies del mismo género. Elegimos especies europeas del grupo *obscura* que presentaban características particularmente interesantes para abordar el estudio evolutivo de la región del genoma que comprende el gen *Adh: D. subobscura, D. guanche, D. madeirensis* y *D. ambigua*.

El estudio de las homologías cromosómicas en las relaciones sinápticas o apareamiento de los cromosomas homólogos de los híbridos interespecíficos y el análisis de los polimorfimos enzimáticos, realizados por Costas Krimbas, del Colegio de Agricultura de la Universidad de Atenas, en 1984 y por Montserrat Papaceit, de nuestro departamento, en el mismo año confirman la proximidad genética, especialmente entre *D. suboscura, D. guanche* y *D. madeirensis*. Estas dos últimas especies son endémicas de las islas Madeira y Canarias, respectivamente, y acaban de ser descritas por María Monclús, de nuestro departamento. Se ha hipotetizado que estas especies serían descendientes de las ancestrales que dieron lugar a las líneas de *obscura* y *melanogaster*, ya que también comparten con *D. melanogaster* bastantes rasgos morfológicos. *D. ambigua* es una especie morfológicamente muy parecida a *D. subobscura*, pero desde el punto de vista cromosómico y enzimático se encuentra muy próxima a las especies americanas del mismo grupo.

Había que empezar por cartografiar la región *Adh* mediante hibridación *in situ* de los cromosomas politénicos de las especies de *Drosophila* citadas y comparar, luego, los resultados con los ya establecidos sobre la evolución cromosómica. Los cromosomas politénicos son cromosomas gigantes formados por la yuxtaposición de cadenas de ADN de cromosomas reduplicados que no alcanzan a separarse, lo que facilita su observación en el microscopio óptico, acceso que no suele darse en el caso de los cromosomas interfásicos. Los politénicos se encuentran en determinados tejidos de las larvas de algunos insectos, entre ellos *Drosophila*. Las regiones condensadas aparecen en forma de bandas, mientras que las desespiralizadas se visualizan más tenues y reciben el nombre de interbandas.

Una de las técnicas que se utiliza para asignar una banda de un cromosoma politénico a un gen estructural consiste en desnaturalizar el ADN cromosómico de una preparación citológica y ponerlo en condiciones tales que pueda formar híbridos con un frag-

mento de ADN marcado, homólogo del gen en cuestión. Este fragmento específico, o sonda, puede ser el ADN de un gen ya aislado de una especie próxima. En nuestro caso, marcamos el plásmido *s*AC1 con nucleótidos biotinados; lo que conseguimos provocando roturas en el ADN bicatenario circular con ADNasa I en presencia de dUTP-biotinado, un precursor de la uridina monofosfato, y los 4 nucleósidos trifosfatos usuales: dCTP, dATP, dGTP, dTTP. La ADN pol I, enzima que es capaz de copiar una cadena modelo de ADN incorporando los nucleótidos complementarios, alarga la discontinuidad que ha producido la ADNasa I y dirige la incorporación de dUMP-biotinado a medida que copia la cadena intacta. Esta reacción ("nick-translation", en inglés) sólo difiere de la convencional en que se utiliza un nucleótido biotinado en vez de un nucleótido radiactivo.

La sonda biotinada se pone en presencia de los cromosomas politénicos desnaturalizados mediante un tratamiento con solución alcalina y temperatura de 80 grados C. Para la detección de los híbridos "ADN cromosómico: sonda biotinada" se utilizó un producto comercial: el complejo Detek I-hrp. Este complejo está formado por estreptavidina, proteína que presenta una gran afinidad por la biotina, y peroxidasa biotinada de rábano. La estreptavidina se enlaza por un lado con los nucleótidos biotinados de la sonda y, por otro, con la peroxidasa de rábano. Al introducir en el medio un substrato de la peroxidasa y un colorante para detectar la reacción enzimática, aparece un color marrón oscuro sobre la banda que ha hibridado con la sonda, localizándose así el locus *Adh*.

La localización del gen *Adh* por hibridación *in situ* sobre los cromosomas politénicos de *Drosophila* de las especies *D. subobscura*, *D. madeirensis* y *D. ambigua* permite obtener puntos de referencia con los que comparar los modelos de bandas adyacentes a dicha señal, establecer la homología entre cromosomas de distintas especies y determinar las reordenaciones que han tenido lugar durante la evolución. Gracias a esta técnica se ha establecido ya la homología entre los elementos cromosómicos de *D. melanogaster, D. subobscura, D. madeirensis* y *D. ambigua*. Se sabe que son homólogos el brazo izquierdo (*L*) del cromosoma 2 (metacéntrico) de *D. melanogaster*, el cromosoma *U* (acrocéntrico) de *D. subobscura*, el cromosoma *U* (acrocéntrico) de *D. madeirensis* y el cromosoma *E* (metacéntrico) de *D. ambigua*. Al comparar la posición del gen *Adh* en las distintas especies se observa que en *D. melanogaster* se halla en la mitad más cercana al centrómero, que en *D. subobscura* y *D. madeirensis* está adyacente al centrómero y que en *D. ambigua* se ubica en la mitad del brazo derecho más alejada del centrómero. De ello se deduce que a lo largo de la evolución se produjeron distintas inversiones pericéntricas y paracéntricas que justifican las diferentes localizaciones del gen *Adh* en estas especies. De momento, no es posible, sin embargo, asignar un orden temporal a estos sucesos.

Abordamos también el estudio del gen *Adh* desde otra óptica: el análisis de las dianas de restricción en esta región genómica. (Se llama diana de restricción a la secuencia específica de ADN reconocida por una enzima de restricción.) Purificamos el ADN de las cuatro especies mencionadas y también el de *D. melanogaster*, que utilizamos como referencia. Después de digerir el ADN con una enzima de restricción, *Eco* RI o *Hind* III, separamos los fragmentos por electroforesis y los transferimos a filtros de nitrocelulosa (a través de la llamada técnica de Southern). La hibridación de los fragmentos con la sonda *s*AC1, esta vez marcada con nucleótidos radiactivos, indicaba cuáles eran los fragmentos de ADN que contenían una secuencia homóloga. Las dianas de la enzima *Eco* RI se mantenían en todas las especies, mientras que se observaban algunas diferencias en el caso de *Hind* III. Estos resultados están de acuerdo con la proximidad genética de estas especies, reflejada también en la región *Adh*, si bien es necesario analizar muchas más dianas de restricción y determinar la secuencia de nucleótidos, para conocer con detalle evolución.

Los resultados obtenidos sobre el sistema ADH de *Drosophila* han ayudado a comprender algunos aspectos evolutivos y adaptativos de interés. Sin embargo, queda aún un largo camino por recorrer antes de que podamos explicar las causas moleculares de la adaptación biológica y los mecanismos de control de la expresión génica, en beneficio del conocimiento de algunos de los procesos básicos de la biología molecular y del conocimiento científico en general.

El teorema enorme

La clasificación de los grupos finitos simples carece de precedentes en la historia de la matemática, pues su demostración ocupa 15.000 páginas. Lo exótico de la solución ha despertado inusitado interés en otros campos muy alejados

Daniel Gorenstein

¿Cómo es posible que, para demostrar un solo teorema matemático, sean precisas 15.000 páginas? ¿Quién será capaz de leer tal demostración y quién podrá emitir un juicio acerca de su validez? Y no obstante, ahí está: la demostración de que se han hallado todos los grupos finitos simples ha terminado por ocupar unas diez o quince mil páginas. Como es obvio, la responsabilidad de semejante hazaña no es atribuible a una persona concreta; tampoco puede achacarse la longitud de la demostración a la necesidad de laboriosos cálculos con ordenador (aunque sí se utilizaron ordenadores en cierta etapa del análisis). La obra, por el contrario, es fruto de los esfuerzos conjuntos de más de 100 matemáticos, en su mayoría estadounidenses, ingleses y alemanes, pero también canadienses, australianos y japoneses. La demostración completa se encuentra dispersa por las páginas de más de 500 artículos, publicados casi todos entre finales del decenio de 1940 y comienzos del actual en numerosas revistas técnicas.

La noción de grupo matemático ha sido fundamental en la matemática desde que fue introducida, a comienzos de la década de 1830, por Evariste Galois. Este matemático francés se sirvió de propiedades de los grupos para zanjar, negativamente, una cuestión que ya entonces tenía más de 200 años, a saber: ¿será posible expresar las raíces de las ecuaciones polinómicas de grado quinto, o superior, por medio de fórmulas similares a las conocidas para las ecuaciones cuadráticas (y las no tan conocidas, pero existentes, para ecuaciones de tercer y cuarto grados)? Aunque la aplicación en que Galois estaba pensando se limitaba a las ecuaciones polinómicas, la noción de grupo ha resultado presentarse a través de toda la matemática y en la naturaleza.

Los giros de una esfera, las periodicidades de la estructura de un cristal y las simetrías atómicas son otros tantos ejemplos de grupos. Incluso las interacciones entre las partículas elementales y la "óctuple vía", en virtud de la cual ciertas partículas son descritas como objetos compuestos formados por quarks, requieren ser formuladas con auxilio de nociones de teoría de grupos para resultar comprensibles.

El sistema formado por los números enteros juntamente con la operación de adición es un grupo; y es que, en realidad, las reglas relativas a las combinaciones de elementos de un grupo proceden directamente de algunas de las reglas de la aritmética ordinaria, si bien dándoles forma más general. En vista de lo asombrosamente aplicable que es la aritmética a la vida cotidiana, mal puede sorprendernos que los mismos conceptos, aplicados a situaciones más abstractas, lleguen a convertirse en poderosos instrumentos para la comprensión del universo.

Los grupos simples son los bloques fundamentales con los que están edificados todos los grupos. El término grupo simple no tiene connotaciones estructurales. Exactamente como sucede con los átomos, cuya estructura interna es de enorme complejidad, así les ocurre a los grupos simples. Los grupos simples únicamente son simples en el sentido de que, al igual que los átomos, es imposible descomponerlos en unidades menores de la misma naturaleza.

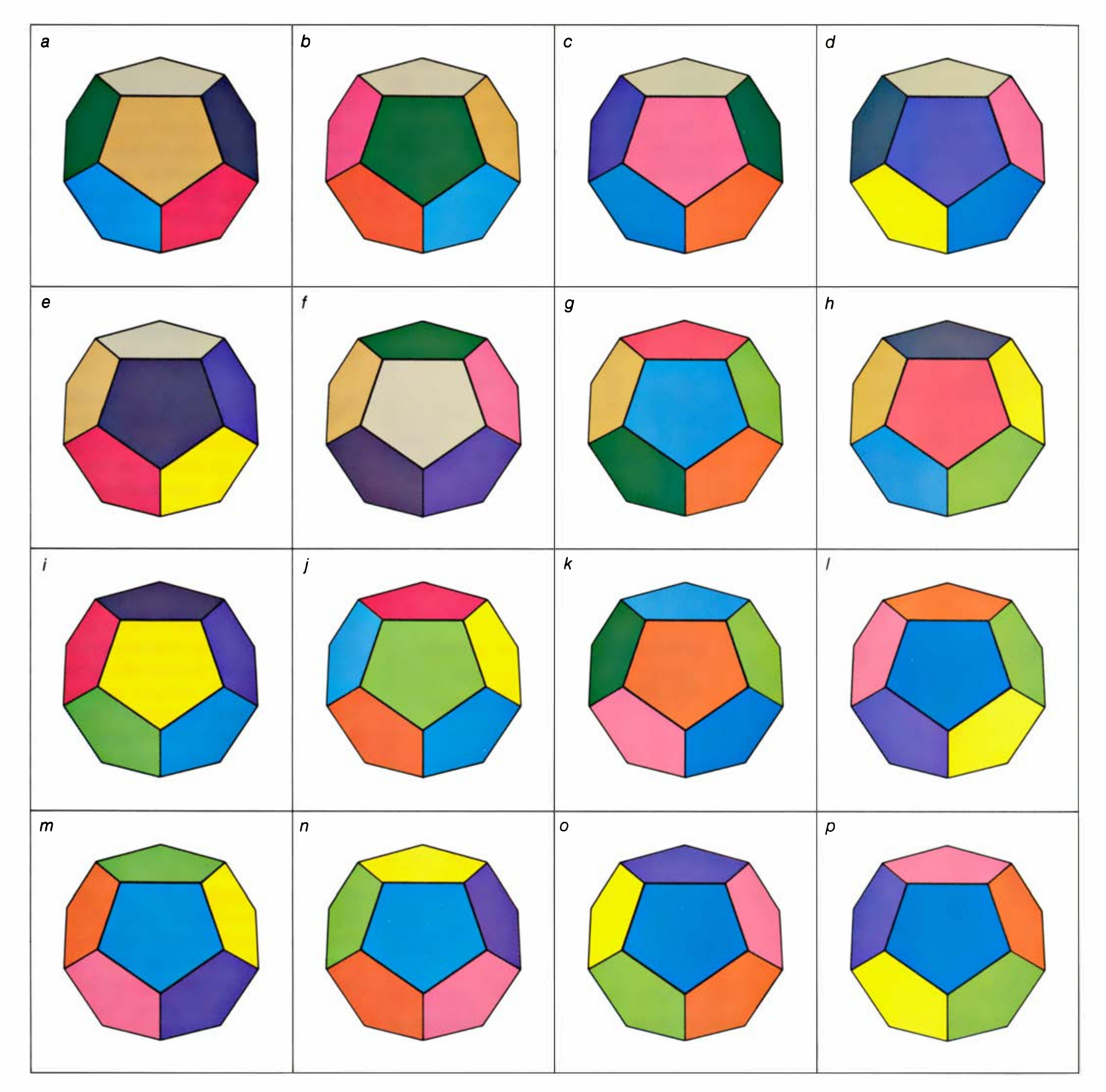
Muchos grupos constan de un número infinito de elementos, o miembros; lo mismo que antes, el sistema de los números enteros nos proporciona un buen ejemplo, pues no existe limitación para el tamaño de los números a sumar. Por otra parte, las reglas de composición de elementos en un grupo son satisfechas, también, por muchos sistemas de tamaño finito. Tales sistemas se denominan grupos finitos. Por ejemplo, en los relojes se suman las horas lo mismo que los números, pero el resultado se expresa como una hora comprendida entre la 1 y las 12. Cinco horas después de las 10 no son las 15, sino las 3. Las 12 horas, juntamente con la regla habitual de adición horaria, forman el grupo de los enteros módulo 12, o más brevemente, el llamado "grupo (cíclico) del reloj", un grupo de 12 elementos. Se pueden construir sin dificultad grupos cíclicos de períodos distintos de 12 [*véase la figura 2*].

Así pues, ¿qué es exactamente un grupo? La definición formal es precisa y rigurosa, pero tal vez sea más fácil de comprender mediante ejemplos concretos. Un grupo es un conjunto de elementos, junto con una operación, aquí denotada por un asterisco (*), que permite combinarlos. Cualesquiera que sean los elementos a y b del conjunto, tiene que estar definido el "producto" $a*b$, que ha de ser un elemento del conjunto. Por poseer esta propiedad respecto del producto, se dice que el conjunto es "cerrado" respecto de la operación y que la operación es una ley de composición interna. Para que el sistema sea grupo, se exige que la operación cumpla tres reglas. Primera, el conjunto debe contener un "elemento neutro" (también llamado elemento idéntico o "identidad" del grupo), denotado e, tal que $a*e$ y $e*a$ sean ambos iguales a a. Segunda, para cada elemento a del conjunto tiene que existir en el conjunto un elemento llamado inverso de a, que se denota a^{-1}, tal que los productos $a*a^{-1}$ y $a^{-1}*a$ sean iguales a e. Finalmente, la operación * tiene que ser asociativa. Dicho de otro modo, cualesquiera que sean los tres elementos a, b y c que se tomen en el conjunto, la secuencia en que se efectúe la operación * no afecta al resultado: $a * (b * c) = (a * b) * c$.

Fijémonos en que la definición no exige que la operación * haya de ser conmutativa. Es decir, que, en general, $a * b$ y $b * a$ no tienen por qué ser igua-

les. En realidad, la mayor parte de los grupos interesantes que se presentan, tanto en la matemática como en la naturaleza, no son conmutativos [*véase la figura 6*].

Podemos apreciar ahora por qué las reglas de combinación de los elementos de un grupo son las leyes formales básicas de la aritmética, si bien enunciadas de forma más abstracta. Examinemos para ello cómo los enteros forman grupo cuando la operación para combinarlos es la adición ordinaria. Lo que en el grupo hemos llamado producto es ahora la suma habitual; por ejemplo, 5*7 es 5+7, o sea, 12. Además, el elemento neutro es el número 0; el "inverso" del número a es el número $-a$; y para cualesquiera números a, b y c se verifica la propiedad asociativa: por ejemplo, 3+(4+5) es igual a (3+4)+5. Análogamente, el conjunto de los números racionales no nulos, con la operación de multiplicación ordinaria, forma también un grupo: el número 1 es el elemento neutro, y el número $1/a$ es el inverso al número a. (Es preciso



1. LAS OPERACIONES DE SIMETRIA del dodecaedro regular forman un objeto matemático llamado grupo. Los grupos han sido una de las materias de mayor interés para los matemáticos a lo largo de los 150 últimos años. La teoría de grupos ha encontrado también importantes aplicaciones extra-matemáticas, especialmente en cristalografía y en la física de las partículas elementales. Todos los grupos finitos están construidos a partir de los llamados grupos simples, que desempeñan en la teoría de grupos un papel similar al de los átomos en la física, o al de los números primos en la aritmética. Los giros del dodecaedro que conservan su orientación en el espacio forman el más pequeño de todos los grupos simples cuyos elementos no conmutan todos unos con otros; dicho de otro modo, la posición final del dodecaedro con respecto a una dirección prefijada del espacio depende del orden en que se lleven a cabo los giros. El grupo consta de 60 elementos, que corresponden a las cinco posiciones de cada una de las 12 caras de este poliedro; en la ilustración vemos cada una de las caras en posición frontal (*a–l*) y las cinco posiciones frontales de la cara azul oscura (*l–p*). El grupo ha desempeñado un papel destacado en la historia de la matemática. Evariste Galois, inventor de la teoría de grupos, demostró que la simplicidad de este grupo, más el hecho de que su número de elementos no sea primo, permite elucidar un problema clásico: la ecuación polinómica general de coeficientes racionales, cuyo término de grado máximo sea x^5, no se resuelve con radicales.

excluir el número 0, porque carece de inverso.) Las 12 horas del reloj forman un grupo, en el cual las 0 horas, o si se quiere, las 12 horas, son el elemento neutro; mientras que, pongamos por caso, el elemento inverso de las 5 son las 7. Estos ejemplos de grupos aritméticos son conmutativos, porque ni la adición ni la multiplicación de números dependen del orden en que se opere con éstos. Por otra parte, los grupos cuyos elementos son ciertos giros y ciertas simetrías de figuras geométricas regulares, tales como esferas y triángulos equiláteros, casi siempre resultan ser no conmutativos.

Existe una analogía entre la teoría de grupos y la multiplicación ordinaria de números, que puede ayudarnos a imaginar cómo se forman los grupos complejos a partir de los grupos simples. Cada número entero es, como se sabe, ora primo, ora compuesto. Los números primos, como 2, 3, 5, 7 y 11, tan sólo son exactamente divisibles por sí mismos y por 1. Los números que no son primos son compuestos; el llamado teorema fundamental de la aritmética enuncia que todo número compuesto es descomponible en producto de un único conjunto de números primos. El número compuesto 12, pongamos por caso, tiene los factores primos 2, 2 y 3.

Hay en teoría de grupos un proceso llamado descomposición por homomorfismo que permite escindir un grupo finito arbitrario en un conjunto de grupos simples, unívocamente determinado, de modo muy similar a como un número es descompuesto en factores primos. Además, el número de elementos de cada uno de los grupos simples componentes es divisor del número de elementos del grupo paterno, y el producto de todos esos números es igual al número de elementos del grupo paterno.

Sin embargo, hay que tener cuidado y no llevar demasiado lejos la analogía. En efecto, los grupos simples componentes de un grupo determinado pueden contener un número compuesto de elementos. Y más importante todavía: mientras que el producto de todos los números primos de un conjunto dado (como los números del conjunto 2, 2, 3, 5) es un número compuesto unívocamente determinado (en este caso, 60), por lo general, combinando de distintas formas los grupos simples de una colección dada, se pueden formar muchos grupos no equivalentes.

No es solamente la extraordinaria longitud de la demostración lo que hace a la clasificación de los grupos finitos simples tan insólita en los anales de la matemática, sino también la desconcertante naturaleza de la solución. Conforme iba progresando la investigación en el problema, los especialistas en teoría de grupos iban descubriendo no sólo familias infinitas de grupos simples (al final se habían construido 18 de estas familias), sino también cierto número de grupos sumamente irregulares, imposibles de encajar en ninguna de las 18 familias regulares. Los cinco primeros de estos paradójicos grupos simples esporádicos (como se dio en llamarlos) habían sido descubiertos por Emile Mathieu, en el decenio de 1860; el menor de los grupos de Mathieu contiene exactamente 8 × 9 × 10 × 11 = 7920 elementos. El sexto grupo esporádico, que consta de 175.560 elementos, fue descubierto por Zvonimir Janko, a la sazón en la Monash University, en Australia, un siglo entero después de los trabajos de Mathieu.

A partir de ese momento, las nuevas y extrañas criaturas comenzaron a surgir súbitamente, más o menos, a razón de una por año, marchando al paso del intenso desarrollo teórico de los dos últimos decenios. La agitación que produjeron estos descubrimientos salpicó a la gran comunidad matemática. El punto culminante se alcanzó en 1982, cuando Robert Griess, Jr., que se encontraba entonces en el Institute for

CONJUNTO	OPERACION DEL GRUPO	ELEMENTO NEUTRO	ELEMENTO INVERSO
ENTEROS ORDINARIOS 0, ±1, ±2,...	ADICION [2 + 3 = 5; 3 + (−7) = −4]	0 [3 + 0 = 0]	−a [3 + (−3) = 0]
NUMEROS RACIONALES p/q NO NULOS	MULTIPLICACION $\left[\frac{2}{3} \times \frac{4}{5} = \frac{8}{15}\right]$	1 $\left[\frac{2}{3} \times 1 = \frac{2}{3}\right]$	$\frac{1}{p/q}$ $\left[\frac{2}{3} \times \frac{1}{2/3} = 1\right]$
LAS 12 HORAS DEL RELOJ (12 or 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11)	ADICION MODULO 12 [3 + 4 = 7; 6 + 8 = 2]	LAS 0 HORAS [LAS 3 + LAS 0 = LAS 3]	LAS 12 −a HORAS [LAS 2 + LAS (12−2) = LAS 0 HORAS]
UN RELOJ DE 11 HORAS (11 or 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)	ADICION MODULO 11 [3 + 4 = 7; 6 + 8 = 3]	LAS 0 HORAS [LAS 3 + LAS 0 = LAS 3]	LAS 11−a HORAS [LAS 2 + LAS (11−2) = LAS 0 HORAS]
UN RELOJ DE 3 HORAS (3 or 0, 1, 2)	ADICION MODULO 3 [2 + 1 = 0; 2 + 2 = 1]	LAS 0 HORAS [LAS 2 + LAS 0 = LAS 2]	LAS 3−a HORAS [LAS 2 + LAS (3−2) = LAS 0 HORAS]

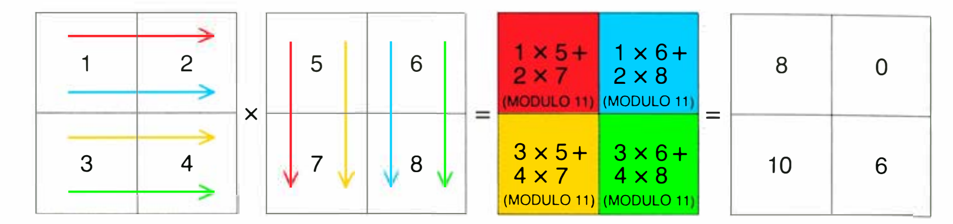
2. VARIOS EJEMPLOS DE GRUPOS se recogen en la tabla. Cada grupo está formado por un conjunto de elementos y una operación de grupo, de ordinario llamada "multiplicación", merced a la cual los elementos del grupo pueden combinarse por pares y producir un elemento del grupo (con frecuencia distinto de los anteriores). La operación del grupo, que podemos denotar mediante un asterisco (∗), se comporta de modo muy similar a la multiplicación o la adición ordinarias. Tiene que cumplir la ley asociativa: para cualesquiera elementos a, b y c que se tomen en el grupo, el "producto" triple $(a*b)*c$ tiene que ser igual al producto triple $a*(b*c)$. El conjunto de elementos ha de incluir un elemento unidad, designado por e, tal que $e*g$ y $g*e$, los productos de cualquier elemento g por el elemento identidad, e, sean ambos iguales a g. Para cada elemento g del conjunto ha de existir también un elemento inverso g^{-1}, tal que $g*g^{-1}$ y $g^{-1}g*$, que son los productos de g por su inverso, sean ambos iguales al elemento identidad, e. El conjunto de elementos de un grupo puede ser infinito, como en los dos primeros ejemplos de la tabla superior, o finito, como en los tres últimos de la misma. Los ejemplos cuarto y quinto son los llamados grupos cíclicos de períodos 11 y 3, respectivamente; sus elementos se suman como si fueran las horas de un reloj de 11 horas, o de tres horas, respectivamente.

Advanced Study, construyó un grupo que vino a ser conocido por "el monstruo", porque el número de elementos que lo compone es de 808.017.424-794.512.875.886.459.904.961.710.757-005.754.368.000.000.000, aproximadamente, 8×10^{53}. Finalmente, se descubrieron 26 grupos simples esporádicos. El monstruo es el mayor de todos ellos pero, en razón de las muchas simetrías internas que presenta, Griess lo ha rebautizado "el gigante amable".

Dar con un grupo simple es una cosa, cuya recompensa es el propio descubrimiento, pero demostrar que se tienen todos es harina de muy otro costal. Y he aquí lo que afirma el teorema de clasificación: el universo de los grupos simples finitos está compuesto por 18 familias de grupos, regulares e infinitas, más los 26 grupos esporádicos, ¡y nada más! La demostración de este enunciado ha sido tema de 500 artículos, que totalizan alrededor de 15.000 páginas.

En el curso normal de acontecimientos, es corriente que se conjeture explícitamente un resultado mucho antes de demostrarlo. Cuando se formula una conjetura es por razones excelentes: o bien el resultado es sugerido por investigaciones anteriores, o bien existen ejemplos matemáticamente no triviales para los cuales la conjetura es cierta. Sin embargo, a lo largo de casi toda la guerra librada contra los grupos finitos simples, no les fue posible a los especialistas estimar siquiera el tamaño de las fuerzas enemigas. Entraba dentro de lo plausible que fueran muchos –muchísimos quizá– los grupos simples esporádicos que permanecieran fortificados e invisibles. Consiguientemente, no fue sino muchos años después de comenzados los trabajos de clasificación cuando los teóricos pudieron plantearse siquiera la formulación del teorema que confiaban en poder algún día demostrar.

Tal realidad obligó a una estrategia cautelosa. Se fueron librando batallas limitadas tan sólo a tipos especiales de grupos simples. Estas victorias parciales, duramente logradas, fueron proporcionando teoremas de clasificación de tipos restringidos, que gradualmente desenterraron y desarraigaron los grupos esporádicos restantes. El proceso queda bien ilustrado por el descubrimiento del sexto grupo esporádico (el grupo encontrado por Janko), un grupo emparentado con la decimoséptima familia de grupos simples, hallada en 1960 por Rimhak Ree, de la Universidad de Columbia Británica.



3. "MULTIPLICACION" de dos matrices de dimensión dos por dos; se efectúa de acuerdo con el procedimiento aquí mostrado. Existen reglas similares para multiplicar dos matrices *n* por *n*, para cualquier número *n*. La matriz producto es una matriz cuadrada que tiene el mismo número de elementos que las dos matrices factores. Los elementos de las matrices pueden tomarse de entre los elementos de un sistema numérico cualquiera (o cuerpo), que es un sistema matemático cuyos elementos pueden ser sumados, restados, multiplicados y divididos de modo muy similar a como se hace con los números ordinarios. Los números racionales y los números complejos son ejemplos familiares de sistemas numéricos infinitos, y los elementos de los grupos cíclicos de período primo proporcionan ejemplos de algunos de los posibles sistemas numéricos finitos. Aquí, los elementos pertenecen al grupo cíclico de período 11. Para calcular en este sistema numérico finito tenemos que calcular primero el resultado según la aritmética ordinaria y hallar después el resto de la división entre 11. Por ejemplo, el elemento $1 \times 5 + 2 \times 7$ del ángulo superior izquierdo del producto de matrices (la tercera matriz desde la izquierda) es igual a 19, y el resto, tras dividirlo por 11, es 8. Las matrices suelen considerarse objetos matemáticos propios. Todos los grupos finitos y, en particular, todos los grupos finitos simples, pueden representarse mediante grupos de matrices combinadas por multiplicación.

En el seno de todo grupo simple se encuentran subgrupos menores, llamados centralizadores de involuciones, que desempeñan importante papel para comprender la estructura del grupo paterno. En el caso de los grupos de Ree, los centralizadores de involuciones pueden ser representados mediante matrices cuadradas de 2×2, cuyos cuatro elementos se extraen de un sistema finito de números de tamaño igual a una potencia de 3, con exponente impar. Por ejemplo, si dicho exponente impar fuera 1, el sistema finito de números estaría compuesto por los tres elementos del grupo de clases residuales de enteros, módulo 3, llamado a veces grupo del reloj (o más técnicamente, grupo cíclico) de período 3. Para poder demostrar uno de los primeros teoremas restringidos de clasificación, era necesario probar que los grupos de Ree son los únicos grupos simples que gozan de la siguiente propiedad: sus centralizadores de involuciones pueden ser representados mediante matrices de 2×2 cuyos elementos pertenezcan a un sistema numérico finito de tamaño igual a una cierta potencia impar de un cierto número primo p. Se trataba de un primer paso, muy natural, tendente a demostrar la conjetura siguiente: si un grupo simple arbitrario tiene la propiedad recién mencionada, el tamaño del sistema numérico finito al que han de pertenecer los elementos de las matrices 2×2 ha de ser una potencia impar del número primo 3. Con el tiempo, esta conjetura pudo ser confirmada, a excepción del único caso en que p es igual a 5, y el exponente de la potencia impar es igual a 1.

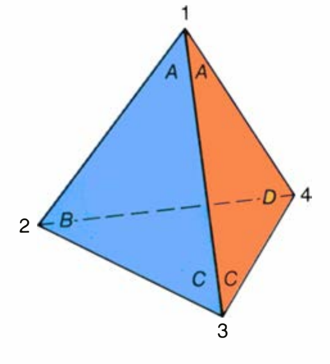
Janko se puso a estudiar el caso excepcional del grupo de Ree, plenamente confiado en que no habría grupos simples del tipo dado que tuvieran un sistema numérico de tamaño 5^1, o sea, 5. A pesar de todos sus esfuerzos,

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 8 & 2 & 10 & 10 & 8 & 10 & 8 \\ 9 & 1 & 1 & 3 & 1 & 3 & 3 \\ 10 & 10 & 8 & 10 & 8 & 8 & 2 \\ 10 & 8 & 10 & 8 & 8 & 2 & 10 \\ 8 & 10 & 8 & 8 & 2 & 10 & 10 \\ 1 & 3 & 3 & 9 & 1 & 1 & 3 \\ 3 & 3 & 9 & 1 & 1 & 3 & 1 \end{pmatrix}$$

4. DOS MATRICES DE SIETE POR SIETE generan la totalidad de las matrices que componen los elementos del grupo simple esporádico descubierto en 1965 por Zvonimir Janko. Los elementos de las matrices se toman de entre los elementos 0 a 10 del grupo cíclico de período 11. El grupo de Janko fue el sexto grupo simple esporádico descubierto, y el primero descubierto en más de 100 años. Está formado por todos los posibles productos matriciales de las dos matrices generadoras, por ejemplo *AA*, *BB*, *ABA*, *BBAABABBB*, y así sucesivamente. El grupo consta de $2^3 \times 3 \times 5 \times 7 \times 11 \times 19$, o sea, 175.560 matrices.

no pudo eliminar esta posibilidad, y así terminar la demostración de la conjetura. Por el contrario, con trabajo considerable logró demostrar que si tal grupo existiese, tendría que contener exactamente $2^3 \times 3 \times 5 \times 7 \times 11 \times 19 = 175.560$ elementos.

A Janko le resultaba demasiado difícil creer que fuese posible establecer una conclusión tan taxativa sin que existiera un auténtico grupo acechando en el trasfondo. Persistió en su empeño con renovada expectación, sosteniendo que si tal grupo se diese tendría que estar engendrado por dos matrices de siete por siete, cuyos elementos de filas y columnas habrían de tomarse del grupo de clases de restos módulo 11 [*véase la figura 4*]. Dicho de otro modo, si las dos matrices generadoras fuesen llamadas A y B, el grupo estaría com-



5. LOS GIROS DE UN TETRAEDRO REGULAR que conservan la orientación del poliedro en el espacio forman un grupo de 12 elementos. Los giros se combinan secuencialmente; es decir: el producto de dos giros es el resultado de efectuar un giro, e inmediatamente, el siguiente. Los cuatro giros de la primera hilera son el giro "identidad", que no hace nada, y tres giros "de arista". Cada uno de los giros de arista intercambia en el tetraedro dos pares de vértices del tetraedro; por ejemplo, el giro (12)(34) intercambia los vértices situados en las posiciones 1 y 2 y los vértices de posiciones 3 y 4. La hilera central muestra el resultado de componer cada uno de los giros de arista con el giro "de cara" (123), que envía el vértice situado en 1 a la posición 2, el vértice de la posición 2 a la posición 3, y el vértice situado en la posición 3 de nuevo a la posición 1. La última hilera muestra el producto de cada giro de arista con el giro de cara (132). Cada producto equivale a un giro de cara. Los giros son aplicados al tetraedro ilustrado abajo de todo, a la izquierda.

puesto por todos los posibles productos matriciales de estas dos matrices, repetidas un número finito de veces: *AA, BB, ABA, BBAABABBB*, y así sucesivamente.

La única cuestión pendiente era verificar si el grupo generado por estas dos matrices constaba de 175.560 elementos exactamente. De no ser así, el análisis de Janko hubiera proporcionado la contradicción que había buscado desde el principio. A primera vista, pudiera parecer notable que todos los posibles productos de A y B, repetidos, hayan de ser equivalentes a una de las matrices de un conjunto que tan sólo consta de 175.560 matrices, pues eso es precisamente lo que supone la existencia de un grupo del tipo requerido. Por ejemplo, tendrían que pertenecer al conjunto productos de más de un millón de A y B. En realidad, hay 11^{49}, o sea, aproximadamente 10^{51}, matrices de 7×7 cuyos elementos pertenecen al grupo de clases de restos módulo 11; por consiguiente, los productos engendrados a partir de las dos matrices generadoras tan sólo constituyen una diminuta fracción del total de matrices posibles. A pesar de todo, los cálculos, realizados enteramente a mano, acreditaron la existencia del sexto grupo esporádico, hoy llamado J_1, en honor de Janko.

La construcción de J_1, que Janko llevó a cabo, guarda cierto paralelismo con la física de las partículas elementales. El análisis teórico proporcionó indicios de la existencia de un nuevo grupo simple; después, un cálculo explícito, similar a una comprobación experimental, demostró la existencia real de tal grupo.

Este proceso en dos fases queda aún más vigorosamente ilustrado por el segundo de los éxitos de Janko. Si J_1 fue descubierto "disparatando" con las propiedades de los centralizadores de involuciones de los grupos de Ree, ¿por qué no ensayar lo mismo con algunos otros grupos conocidos? Janko cantó bingo casi inmediatamente; en realidad, acertó por partida doble. A partir de un único centralizador de una involución encontró pruebas de la posible existencia de nuevos grupos simples, uno de 604.800 elementos y de 50.232.960 el otro. Esta vez, sin embargo, Janko no consiguió proporcionar la comprobación experimental. Valiéndose de la información que Janko había deducido, Marshall Hall, Jr., y David Wales, ambos del Instituto de Tecnología de California, construyeron el menor de ambos grupos, el J_2, y Graham Higman, de la Universidad de Oxford, y John MacKay, de la Concordia University, en Canadá, el mayor de los dos, J_3. Los cálculos requeridos para J_3 eran demasiado complejos para ser realizados a mano, por lo que fueron efectuados en ordenador.

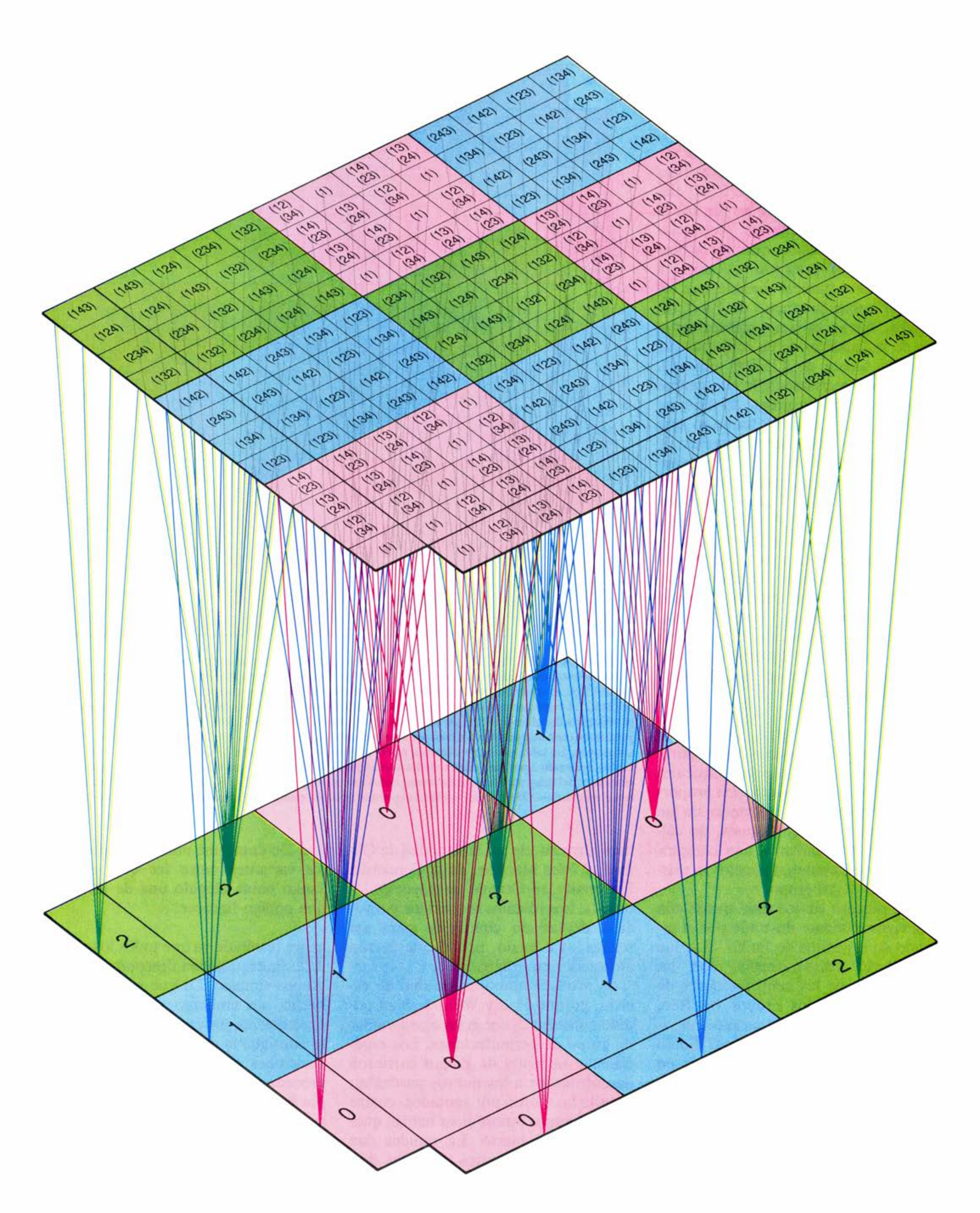
Se produce ahora un cambio de escena, porque Hall y Wales habían podido construir J_2 como un tipo familiar de grupo de permutaciones. Los especialistas en teoría de grupos corrieron desenfrenados a los nuevos pastizales; cuando los dieron por agotados, cuatro nuevos grupos esporádicos habían quedado al descubierto. La rapidez con que se realizaron estos primeros descubrimientos sirvió solamente para reforzar la impresión de que el número de grupos esporádicos aún no descubiertos era grande; infinito quizás. En retrospectiva, sin embargo, los matemáticos se han dado cuenta de que los descubrimientos de J_2 y J_3, también inspirados por Janko, fueron casi milagrosos. De entre una miríada de potenciales variantes entre las que escoger, Janko había elegido una de las pocas que podían florecer.

He resaltado ya que todos los grupos finitos pueden descomponerse en grupos simples. Lo que todavía no he hecho, sin embargo, es decir qué es exactamente un grupo simple, ni cómo se consigue la descomposición. Ambas cuestiones están relacionadas con un procedimiento matemático conocido por homomorfía, al que hice breve alusión antes. Cuando se forma la imagen homomórfica de un grupo, la multiplicación del grupo paterno G queda reflejada en el grupo imagen G', si bien, de ordinario, en forma reducida. Podemos comparar el proceso al de observar un objeto a través de un telescopio: se conservan los rasgos generales del objeto, pero su tamaño aparente queda reducido. [Estas correspondencias "telescópicas" se llaman en la jerga

SEGUNDO ELEMENTO (columnas) — PRIMER ELEMENTO (filas)

*	(1)	(12)(34)	(13)(24)	(14)(23)	(123)	(134)	(243)	(142)	(132)	(234)	(124)	(143)
(1)	(1)	(12)(34)	(13)(24)	(14)(24)	(123)	(134)	(243)	(142)	(132)	(234)	(124)	(143)
(12)(34)	(12)(34)	(1)	(14)(23)	(13)(24)	(134)	(123)	(142)	(243)	(234)	(132)	(143)	(124)
(13)(24)	(13)(24)	(14)(23)	(1)	(12)(34)	(243)	(142)	(123)	(134)	(124)	(143)	(132)	(234)
(14)(23)	(14)(23)	(13)(24)	(12)(34)	(1)	(142)	(243)	(134)	(123)	(143)	(124)	(234)	(132)
(123)	(123)	(243)	(142)	(134)	(132)	(124)	(143)	(234)	(1)	(13)(24)	(14)(23)	(12)(34)
(134)	(134)	(142)	(243)	(123)	(234)	(143)	(124)	(132)	(12)(34)	(14)(23)	(13)(24)	(1)
(243)	(243)	(123)	(134)	(142)	(124)	(132)	(234)	(143)	(13)(24)	(1)	(12)(34)	(14)(23)
(142)	(142)	(134)	(123)	(243)	(143)	(234)	(132)	(124)	(14)(23)	(12)(34)	(1)	(13)(24)
(132)	(132)	(143)	(234)	(124)	(1)	(14)(23)	(12)(34)	(13)(24)	(123)	(142)	(134)	(243)
(234)	(234)	(124)	(132)	(143)	(12)(34)	(13)(24)	(1)	(14)(23)	(134)	(243)	(123)	(142)
(124)	(124)	(234)	(143)	(132)	(13)(24)	(12)(34)	(14)(23)	(1)	(243)	(134)	(142)	(123)
(143)	(143)	(132)	(124)	(234)	(14)(23)	(1)	(13)(24)	(12)(34)	(142)	(123)	(243)	(134)

6. TABLA DE MULTIPLICAR correspondiente al grupo de giros del tetraedro que vimos en la figura 5. Fijémonos en que el grupo contiene elementos que no conmutan. Por ejemplo, el giro (12)(34), seguido por el giro (123), es equivalente al giro (134), pero si se invierte el orden en que se realizan los giros su producto es (243). La región encarnada de la parte superior izquierda de la tabla demuestra que, en el tetraedro, el producto de dos giros de arista es, o bien un giro de arista, o bien el giro identidad. Los tres giros de arista, juntamente con la identidad, forman por sí solos un grupo. Tales grupos, que forman parte de otro, se llaman subgrupos. La tabla puede dividirse en nueve regiones cuadradas (*código de colores*). En el seno de cada región coloreada aparecen cuatro de los elementos del grupo: en los recuadros encarnados aparecen los elementos del subgrupo de giros de arista; en los azules, los productos de los cuatro elementos "encarnados" con el giro facial (123) y, en los verdes, los productos de los elementos "encarnados" con el giro facial (132).

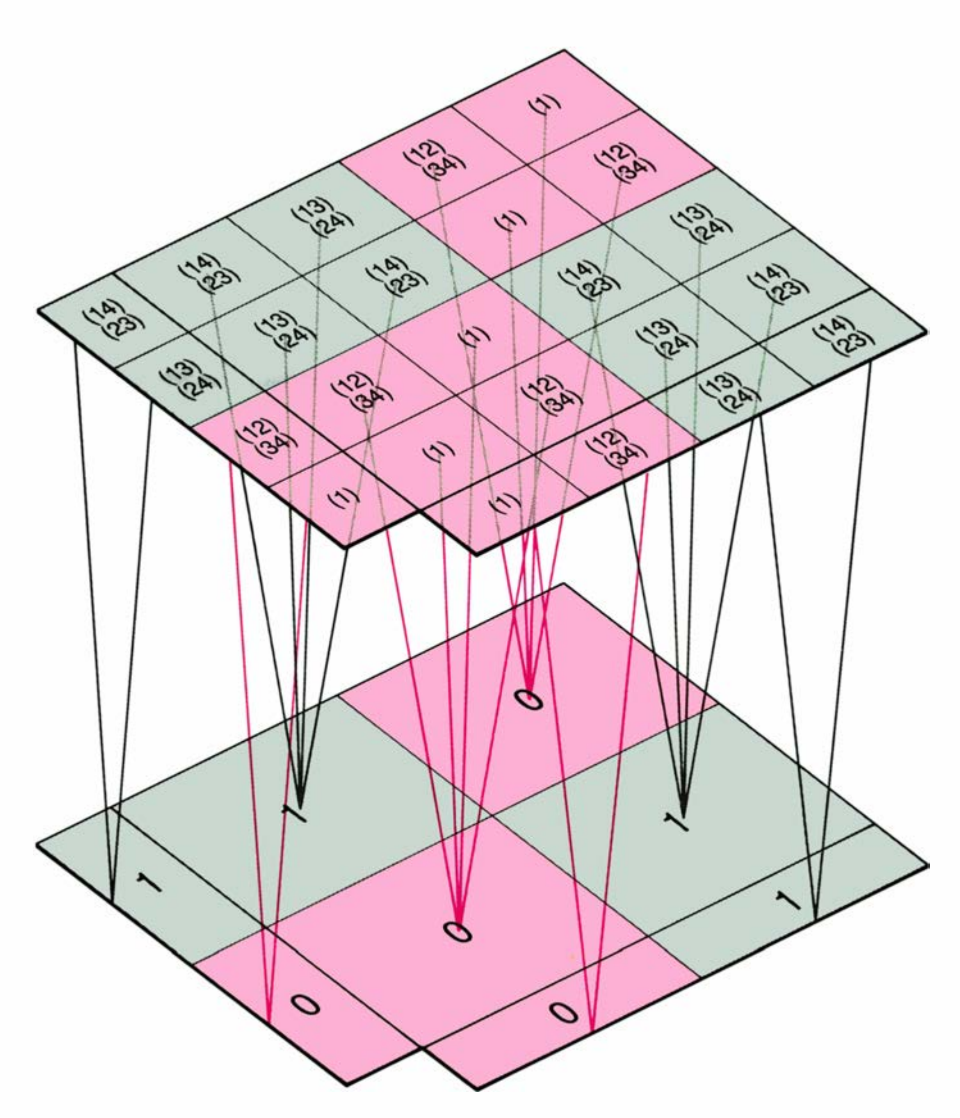


7. LOS GRUPOS SIMPLES COMPONENTES del grupo del tetraedro se obtienen hallando una "imagen homomórfica" del grupo del tetraedro. Cada uno de los elementos coloreados de la tabla de multiplicación del grupo del tetraedro es asociado con el elemento neutro, 0, del grupo cíclico de período 3. Cada uno de los elementos que en la tabla fueron teñidos de azul está asociado con el elemento 1 del grupo cíclico, y cada elemento verde con el 2 del grupo cíclico. La ilustración muestra el modo en que la tabla correspondiente a la adición del grupo es reflejo de la tabla de multiplicar del grupo del tetraedro. Por ejemplo, el producto de dos giros "azules" cualesquiera equivale a un giro "verde", igual que la suma de 1 más 1, los dos elementos "azules" del grupo cíclico, es 2, el elemento "verde" del grupo cíclico. La imagen homomórfica de todo elemento del subgrupo de las rotaciones de arista es 0: el elemento neutro del grupo cíclico.

técnica "epimorfismos" u homomorfismos sobreyectivos. N. del T.]

En el proceso de la correspondencia homomórfica, a cada elemento a del grupo G hay que asociarle un elemento a' del grupo G'; a tal elemento a' se le llama imagen de a. Es perfectamente posible que haya varios elementos de G que tengan la misma imagen en G', lo cual explica la reducción de tamaño. Por otra parte, todo elemento de G' tiene que ser imagen de al menos un elemento de G. Para que un tal método de asociación de los elementos de un grupo G con los elementos de un grupo G' merezca el nombre de correspondencia homomórfica, es preciso que las operaciones internas de G y de G' estén íntimamente emparentadas: si los elementos a' y b' del grupo G' son las imágenes de los elementos a y b del grupo G, el producto $a' * b'$ de G' tiene que ser la imagen del producto $a * b$ en G.

Todo grupo arbitrario G tiene siempre, cuando menos, dos imágenes epimórficas. Una de ellas es su copia idéntica: G' es sencillamente el mismo grupo que G, y cada elemento de G es su propia imagen. En efecto, G queda intacto, por lo que se cumple la condición que pesa sobre el producto para que la correspondencia sea homomórfica. La otra imagen epimórfica que tiene todo grupo es su imagen puntual: aquí G' es el grupo que tiene solamente un elemento, a saber, la identidad e', y e' es la imagen de todo elemento de G. Entonces, indeferentemente de cómo sean elegidos en G los elementos a y b, los tres elementos a, b y $a * b$ tendrán todos la misma imagen e' en G'. En vista de que en G', $e' * e'$ es igual a e', vuelve a cumplirse la condición de la correspondencia homomórfica.



8. LA IMAGEN HOMOMORFICA del subgrupo de los giros de arista da el segundo de los grupos simples que componen el grupo tetraédrico. El giro de arista (12)(34) y el giro identidad (1) forman un segundo subgrupo, incrustado en el anterior (*en color*). Estos giros están asociados con el elemento identidad 0 del grupo cíclico de período 2. Los otros dos elementos del subgrupo de giros de arista, a saber, los giros (13)(24) y (14)(23), están asociados con el elemento 1 del grupo cíclico (*en gris*). La adición módulo 2 del grupo cíclico refleja la multiplicación del subgrupo de giros de arista. Por ejemplo, el producto de un giro "encarnado" y de uno "gris" es equivalente a un giro "gris", exactamente igual que la suma módulo 2 de 0, que es el elemento "encarnado" del grupo cíclico, y de 1, el elemento "gris", es igual a 1, el elemento "gris" del grupo cíclico. Así pues, el grupo cíclico de período 2 es también un grupo componente simple. Un segundo grupo cíclico de período 2 resulta ser el tercero de tales grupos componentes (no mostrado en la ilustración).

Podemos dar ahora la definición de grupo simple. Trátase de un grupo cuyas únicas imágenes epimórficas son sus imágenes idéntica y puntual. Los ejemplos más elementales de grupos simples son los grupos cíclicos de clases de restos módulo un número primo p (que constan precisamente de p elementos). En efecto, un teorema elemental de la teoría de grupos establece que el número de elementos de una imagen epimórfica tiene que dividir exactamente al número de elementos del grupo paterno. Dado que los números primos no tienen más divisores que el número 1 y sí mismos, toda imagen epimórfica de un grupo cíclico de período p tan sólo puede constar, o de p elementos, o de 1 elemento. Ello significa que las únicas imágenes epimórficas de tales grupos son, o imágenes puntuales, o copias de sí mismos. Así pues, los grupos cíclicos de período primo son todos grupos similares, y habida cuenta que la serie de números primos es infinita, tales grupos forman la primera de las 18 familias infinitas de grupos simples y finitos.

Puede resultar útil dar un ejemplo de grupo no simple, a fin de esclarecer cómo el proceso de homomorfía descompone un grupo en sus componentes simples. El conjunto de todos los giros de un tetraedro regular que no cambien la orientación espacial del tetraedro forma un grupo. El resultado de dos cualesquiera de tales giros puede ser expresado como otro giro (posiblemente, de un tercer tipo); para todo posible giro existe un giro inverso que devuelve al tetraedro a su posición de partida. El "giro" que no hace nada actúa con el papel de elemento identidad. Consta este grupo de 12 elementos, que corresponden a las 12 posiciones del tetraedro que conservan su orientación espacial: un conjunto de tres posiciones para cada una de las cuatro caras. Los 12 giros del grupo pueden ser designados por la forma en que intercambian los vértices del tetraedro. Por ejemplo, el giro "de aristas" que intercambia el vértice en posición 1 con el vértice de la posición 2, y el vértice de la posición 3 con el vértice de la posi-

ción 4, puede ser expresado en la forma (12)(34); el giro "de cara" que envía el vértice de la posición 1 a la posición 2, el vértice de posición 2 a la 3, y el 3 nuevamente a la 1, puede ser denotado (123) [*véase la figura 5*].

Se puede demostrar ahora que este grupo G de giros del tetraedro regular admite una imagen epimórfica que no es ni su reproducción idéntica, ni su imagen puntual. La existencia de una tal imagen epimórfica llevará consigo que el grupo del tetraedro no sea simple. Como grupo imagen G' se toma el grupo cíclico de período 3. El grupo del tetraedro, G, puede ser entonces aplicado sobre G' de la manera siguiente: el elemento neutro, 0, de G' ha de ser imagen de los cuatro elementos (1), (12)(34), (13)(24) y (14)(23) de G; el elemento 1 de G' ha de ser imagen de los cuatro elementos (123), (134), (243) y (142) de G, y el elemento 2 de G', imagen de los cuatro elementos (132), (234), (124) y (143) de G [*véase la figura 7*].

No son pocas las comprobaciones que es necesario realizar para confirmar que esta asociación de los elementos de uno y otro grupo tiene la propiedad de ser homomorfismo. Fijémonos, por ejemplo, en los elementos (123) y (134). Su producto en el grupo G, (123)*(134), que es el resultado de aplicar al tetraedro primero el giro (123) y seguidamente el (134), es igual al giro (124) [*véase la figura 6*]. Estos tres elementos tienen en G' las imágenes 1, 1 y 2, respectivamente. Dado que en el grupo G' el producto 1*1 de los elementos 1 y 1 es igual a 1 + 1 (módulo 3), que es igual al elemento 2 de G', se verifica la condición de homomorfismo para los elementos (123) y (134). Otro tanto sucede con cualesquiera otros dos elementos. Dado que G contiene 12 elementos, mientras que G' consta solamente de 3, G' no es ni copia idéntica (isomórfico, en terminología técnica) de G, ni es tampoco una imagen puntual de G. Por consiguiente, G no es simple.

GRUPOS SIMPLES MUY SIMILARES A GRUPOS SIMPLES CONOCIDOS

GRUPOS SIMPLES CONOCIDOS

GRUPO SIMPLE DESCONOCIDO

LINEA DE BRAUER, A MEDIO CAMINO

GRUPOS SIMPLES QUE NO SE PARECEN A GRUPOS CONOCIDOS

9. DEMOSTRACION DEL TEOREMA de clasificación para todos los grupos finitossimples. Exigió probar que un grupo simple cualquiera, cuya estructura de subgrupos es desconocida y quizás arbitrariamente compleja, tiene una estructura de subgrupos equivalente a la de uno de los grupos simples conocidos. En la ilustración se representa la estrategia de la demostración, en la cual los grupos simples conocidos son los puntos situados en la frontera de la región abarcada. Hacia el interior de la región, los puntos representan grupos cuyas estructuras de subgrupos van estando progresivamente más alejadas de las estructuras de subgrupos de los grupos simples conocidos. En el centro de la región se encuentra el grupo simple arbitrario, cuya estructura de subgrupos es desconocida. Cada etapa de la demostración del teorema está aquí representada por el desplazamiento hacia la frontera del grupo simple desconocido; es decir, en cada fase se demuestra que la estructura de subgrupos del grupo arbitrario presenta una semejanza cada vez mayor con la de uno de los grupos simples de la frontera. El trabajo de Richard Brauer condujo a un criterio que describe la etapa situada aproximadamente a mitad de camino hacia la frontera. La extraordinaria longitud de la demostración se explica por la necesidad de explorar más de cien caminos para llegar a la frontera.

¿Cómo se hallan, así pues, las componentes simples del grupo de giros G? Una de las componentes es el propio grupo cíclico G' de período 3. En efecto, G' es imagen homomórfica de G, y G' es simple, pues su período, 3, es número primo. El siguiente paso consiste en fijarse en todos los elementos de G cuya imagen sea el elemento idéntico e' de G': los cuatro giros (1), (12)(34), (13)(24) y (14)(23). Por sí solos, estos cuatro elementos forman un grupo; un grupo así, embebido en otro, se llama subgrupo. Este subgrupo consta de los cuatro giros de arista, que traen al frente cada una de las cuatro caras, pero que no hacen girar la cara una vez que ha alcanzado la posición frontal [*véase la figura 5*].

Tampoco el subgrupo de los giros de arista es simple, pues admite una imagen epimórfica distinta de sí mismo y de la imagen puntual, a saber, el grupo cíclico de período 2. Ahora, el elemento neutro del grupo cíclico es imagen de los giros de arista con elemento identidad (1); el segundo elemento del grupo cíclico es la imagen de los giros de arista (12)(34), (13)(24) y (14)(23). Dado que también el grupo cíclico de período 2 es simple, tal grupo es una segunda componente de la descomposición del grupo paterno G, de los giros del tetraedro [*véase la figura 8*].

El último paso en la descomposición de G consiste en fijarse en los dos elementos (12)(34) y (1) del grupo de rotaciones de aristas cuya imagen sea el elemento idéntico del grupo cíclico de período 2. Estos dos elementos forman también un grupo por sí solos, y las únicas imágenes epimórficas de un grupo de dos elementos son la imagen puntual y la imagen especular. Por consiguiente, el subgrupo formado por el giro de arista (12)(34) y el elemento identidad es por sí mismo un grupo simple, equivalente (isomórfico) al grupo cíclico de período 2. Dado que son imposibles nuevos epimorfismos, el grupo G de giros del tetraedro ha de tener 3 componentes simples, a saber,

los grupos cíclicos de períodos 3, 2 y 2. Porque G contiene 12 elementos, y en vista de que 3 por 2 por 2 es 12, este resultado concuerda con mi anterior observación de que el producto de los números de elementos de las componentes simples de un grupo tiene que ser igual al número de elementos de éste.

Es fácil demostrar que los grupos cíclicos de período primo son los únicos grupos simples conmutativos. En efecto, si un grupo conmutativo consta de un número compuesto, n, de elementos, sucede que, para todo número d que divida exactamente a n, el grupo dispondrá de una imagen epimórfica que conste de exactamente d elementos. Dado que el número d no tiene por qué ser ni 1 ni n, tal grupo conmutativo no es simple. Por ejemplo, el grupo de las cuatro rotaciones de aristas del tetraedro es un grupo conmutativo, en el cual $n = 4$. Dado que 2 es divisor de 4, el número d puede ser 2 y, por consiguiente, el grupo dispone de una imagen epimórfica que consta de dos elementos, como acabamos de ver. Por consiguiente, todos los grupos simples, exceptuados los cíclicos, contienen al menos un par de elementos a y b para los cuales $a*b$ no es igual que $b*a$.

El ejemplo más pequeño de grupo simple no conmutativo contiene 60 elementos. Puede ser descrito como el grupo de giros de un dodecaedro regular que no alteren la orientación espacial del dodecaedro. Los 60 elementos de grupo corresponden a las cinco posibles posiciones de cada una de las doce caras del dodecaedro [*véase la figura 1*].

El grupo del dodecaedro está íntimamente relacionado con el grupo de giros del tetraedro. Estructuralmente, el grupo del tetraedro es idéntico al grupo de las llamadas permutaciones pares de cuatro letras. Por ejemplo, si se permutan las letras $ABCD$, dejándolas en la ordenación $CBAD$, la permutación es impar, porque hay en ella tres inversiones de orden: C precede a B, C precede a A, y B precede a A. Por otra parte, $DCBA$ tiene seis inversiones del orden, y es por ello una permutación par de las letras. El grupo del dodecaedro es estructuralmente idéntico al grupo de las permutaciones pares de cinco letras.

Aunque el grupo del tetraedro no sea un grupo simple, todos los grupos de permutaciones pares de cinco o más letras son grupos simples no conmutativos. Tales grupos, llamados grupos alternantes de n letras, constituyen la segunda familia infinita de grupos simples. Incidentalmente, la noción subyacente al trabajo de Galois sobre la teoría de ecuaciones es esta diferencia de comportamiento entre los grupos alternantes correspondientes a 5 o más letras y los correspondientes a 4 o menos. Distinción que explica las nítidas diferencias entre la naturaleza de las soluciones de las ecuaciones polinómicas de grado menor o igual que 4 y la naturaleza de las soluciones de tales ecuaciones cuando el grado es mayor o igual que 5.

La descripción de las restantes familias de grupos simples es mucho más costosa. Todas ellas tienen representaciones en forma de grupos de matrices cuadradas de tamaño adecuado. En algunos casos, los grupos son inicialmente descritos con el concurso de tales matrices, mientras que en otros hace falta considerable trabajo para obtener sus representaciones matriciales. Lo mismo puede decirse de los grupos simples esporádicos, pero en ellos la situación es todavía más complicada: en ciertos casos, la descripción matricial se funda en cálculos computarizados. Muy notablemente, a pesar de ser el monstruo el mayor de todos los grupos esporádicos, Griess pudo hallar enteramente a mano una representación para él. Su construcción está dada por medio de matrices cuadradas de tamaño 196.883 por 196.883, cuyos elementos son números complejos. Se dan en una tabla los nombres y tamaños de todas las familias de grupos esporádicos [*véase la figura 10*].

A primeros de siglo eran muchas las familias de grupos simples no conmutativos que se habían descubierto, y cada uno de sus grupos, lo mismo que los cinco grupos esporádicos hallados por Mathieu, constaba de un número par de elementos. Pronto se enunció una conjetura en apariencia muy natural: todo grupo finito, simple y no conmutativo, conocido o desconocido, habría de constar de un número par de elementos. Sin embargo, no se pudo probar hasta 1962 que la conjetura (muy famosa, a la sazón) era verdadera, como demostraron Walter Feit y John Thompson, por entonces en la Universidad de Chicago. La complejidad de la demostración de esta conjetura, tan fácil de enunciar y comprender, por otra parte, era precursora de la desmesurada longitud de la clasificación completa de los grupos simples. La demostración de Feit y Thompson ocupaba un número entero del *Pacific Journal of Mathematics*, 255 páginas en total. En 1965, Feit y Thompson fueron premiados con el Premio Cole de Algebra en reconocimiento a su labor.

Su teorema puede ser considerado como ejemplo de teorema restringido de clasificación. En tal forma enuncia que los únicos grupos simples que contienen número impar de elementos son los grupos cíclicos cuyo período sea número primo. Análogamente, el objetivo del teorema completo de clasificación es la obtención de una lista de todos los grupos simples finitos, sin restricciones relativas al número de elementos que puedan contener. Como ya hice notar antes, el teorema final enuncia que la lista completa está formada por las 18 familias regulares e infinitas de grupos más los 26 grupos esporádicos.

En vista de que la demostración del teorema de Feit-Thompson exigió más de 250 páginas, no debe sorprender que la demostración de la conclusión, mucho más complicada, del teorema completo de clasificación haya requerido miles de páginas. La longitud de la demostración no es atribuible a la dificultad de la descripción de los grupos simples conocidos. Los grupos cíclicos pueden ser descritos en pocas líneas; ello no obstante, Feit y Thompson necesitaron cientos de páginas para demostrar que no existían otros grupos simples que constasen de número impar de elementos. La longitud del teorema de clasificación ha de tener otros orígenes.

La causa de tal longitud reside en la complejidad de la estructura del conjunto de subgrupos de un grupo finito arbitrario, al compararla con la de los subgrupos de un grupo simple, lo mismo que una molécula compleja tiene una estructura interna mucho más enrevesada que la de un átomo simple. Para apreciar la naturaleza del problema de clasificación, imaginemos que los grupos finitos simples se encuentran esparcidos sobre una región del plano [*véase la figura 9*].

Los especialistas no podían comenzar su demostración partiendo de la hipótesis de que la estructura de subgrupos de todo grupo simple fuera aproximadamente la misma que la de un grupo simple conocido, pues, en esencia, esa es la tesis del teorema que se proponían demostrar. Así pues, se obtiene una imagen más exacta de lo que había que demostrar si se permite que la estructura de subgrupos de los grupos simples tenga el mismo amplio aba-

nico de complejidad que la estructura de todos los grupos finitos. Los puntos yacentes en la frontera de la región representan los grupos simples conocidos, los puntos cercanos a la frontera representan grupos cuya estructura de subgrupos se asemeja de cerca a las estructuras de los grupos correspondientes a los puntos más cercanos situados en la frontera, y los puntos situados más al interior de la región representan grupos cuyas estructuras guardan progresivamente menos parecido con las estructuras de los grupos simples conocidos.

Al comenzar la demostración se nos da un grupo simple totalmente desconocido, cuya estructura interna, por lo que se alcanza a saber, es tan intrincada como la de cualquier grupo no simple. Representemos el grupo dado como un punto situado en medio de la región plana. Para demostrar el teorema es preciso forzar al grupo dado mediante una serie de razonamientos matemáticos a coincidir con uno de los puntos de la frontera. Cada pequeño desplazamiento del punto hacia la frontera ha de entenderse como un refinamiento de la estructura de subgrupos, inicialmente desconocida, del grupo en cuestión. Conforme el punto se desplaza hacia la frontera, la estructura interna del grupo dado comienza a asumir la forma de uno de los grupos simples conocidos.

Por constituir, cada uno de los puntos de la frontera, un posible destino, el análisis exige explorar un gran número de sendas en el viaje que lleva desde el punto central hasta la frontera. En cada bifurcación del camino se abre una nueva senda; las bifurcaciones corresponden a etapas en las cuales la estructura interna del grupo dado puede tener varias formas alternativas. Al cabo, fue preciso explorar unos 100 caminos antes de que la demostración quedara terminada. Así pues, el teorema de clasificación es la suma de alrededor de 100 teoremas individuales, que tomados conjuntamente producen la conclusión deseada.

En los decenios de 1940 y 1950, los

FAMILIAS INFINITAS DE GRUPOS FINITOS SIMPLES

NOMBRE DE LA FAMILIA	NUMERO DE ELEMENTOS
GRUPOS CICLICOS Z_p DE PERIODO PRIMO p	p, SIENDO p UN NUMERO PRIMO CUALQUIERA EJEMPLOS: $\|Z_2\| = 2$; $\|Z_3\| = 3$; $\|Z_5\| = 5$; . . .
GRUPOS ALTERNANTES Alt_n, FORMADOS POR LAS PERMUTACIONES PARES DE n LETRAS, $n > 4$	$1/2 \times 1 \times 2 \times \ldots \times n$ EJEMPLOS: $\|Alt_5\| = 1/2 \times 1 \times 2 \times 3 \times 4 \times 5 = 60$; $\|Alt_6\| = 360$; . . .
GRUPOS DE CHEVALLEY GRUPOS LINEALES DE CHEVALLEY $A_n(q)$, $n > 1$ O $q > 3$	$q^{n(n+1)/2} \times (q^2-1) \times (q^3-1) \times \ldots \times (q^{n+1}-1)/d$, d = M C D $(n+1,q-1)$ EJEMPLOS: $\|A_1(4)\| = 4^{1\times(1+1)/2} \times (4^{1+1}-1)/[\text{M C D } (1+1, 4-1)] = 4 \times 15/1 = 60$; $\|A_1(5)\| = 5 \times (5^2-1)/[\text{M C D } (1+1,5-1)] = 120/2 = 60$; $\|A_2(2)\| = 168$; . . .
GRUPOS SIMPLECTICOS DE CHEVALLEY $C_n(q)$, $n > 2$	$q^{n^2} \times (q^2-1) \times (q^4-1) \times \ldots \times (q^{2n}-1)/d$, d = M C D $(2,q-1)$ EJEMPLO: $\|C_3(2)\| = 2^{3^2} \times (2^2-1) \times (2^4-1) \times (2^{2\times3}-1)/[\text{M C D } (2,2-1)] =$ $= 512 \times 3 \times 15 \times 63/1 = 1.451.520$
GRUPOS ORTOGONALES DE CHEVALLEY $B_n(q)$, $n > 2$ O $q > 2$	$q^{n^2} \times (q^2-1) \times (q^4-1) \times \ldots \times (q^{2n}-1)/d$, d = M C D $(2,q-1)$
$D_n(q)$, $n > 3$	$q^{n(n-1)} \times (q^n-1) \times (q^2-1) \times (q^4-1) \times \ldots \times (q^{2(n-1)}-1)/d$, d = M C D $(4,q^n-1)$
GRUPOS EXCEPCIONALES DE CHEVALLEY $G_2(q)$, $q > 1$	$q^6 \times (q^6-1) \times (q^2-1)$
$F_4(q)$	$q^{24} \times (q^{12}-1) \times (q^8-1) \times (q^6-1) \times (q^2-1)$
$E_6(q)$	$q^{36} \times (q^{12}-1) \times (q^9-1) \times (q^8-1) \times (q^6-1) \times (q^5-1) \times (q^2-1)/d$, d = M C D $(3,q-1)$
$E_7(q)$	$q^{63} \times (q^{18}-1) \times (q^{14}-1) \times (q^{12}-1) \times (q^{10}-1) \times (q^8-1) \times (q^6-1) \times (q^2-1)/d$, d = M C D $(2,q-1)$
$E_8(q)$	$q^{120} \times (q^{30}-1) \times (q^{24}-1) \times (q^{20}-1) \times (q^{18}-1) \times (q^{14}-1) \times (q^{12}-1) \times (q^8-1) \times (q^2-1)$
GRUPOS DE STEINBERG ${}^2A_n(q)$, $n > 2$ O $q > 2$	$q^{n(n+1)/2} \times (q^2-1) \times (q^3+1) \times (q^4-1) \times \ldots \times (q^{n+1}-(-1)^{n+1}/d$, d = M C D $(n+1,q+1)$
${}^2D_n(q)$, $n > 3$	$q^{n(n-1)} \times (q^n+1) \times (q^2-1) \times (q^4-1) \times (q^6-1) \times \ldots \times (q^{2(n-1)}-1)/d$, d = M C D $(4,q^n+1)$
${}^3D_4(q)$	$q^{12} \times (q^8+q^4+1) \times (q^6-1) \times (q^2-1)$
${}^2E_6(q)$	$q^{36} \times (q^{12}-1) \times (q^9+1) \times (q^8-1) \times (q^6-1) \times (q^5+1) \times (q^2-1)/d$, d = M C D $(3,q+1)$
GRUPOS DE SUZUKI ${}^2B_2(q)$, $q = 2^m$, m IMPAR, $m > 1$	$q^2 \times (q^2+1) \times (q-1)$
GRUPOS DE REE ${}^2G_2(q)$, $q = 3^m$, m IMPAR, $m > 1$	$q^3 \times (q^3+1) \times (q-1)$ EJEMPLO: $\|{}^2G_2(3^3)\| = 27^3 \times (27^3+1) \times (27-1) = 10.073.444.472$; . . .
${}^2F_4(q)$, $q = 2^m$, m IMPAR [PARA $q = 2$, ${}^2F_4(2)'$]	$q^{12} \times (q^6+1) \times (q^4-1) \times (q^3+1) \times (q-1)$ $[\|{}^2F_4(2)'\| = 1/2\|{}^2F_4(2)\| = 1/2 \times 2^{12} \times (2^6+1) \times (2^4-1) \times (2^3+1) \times (2-1) = 17.971.200]$

10. LISTA COMPLETA de todos los grupos finitos simples, junto con el número de elementos de cada grupo. La tabla de esta página da la lista de las 18 familias infinitas de grupos simples, mientras que la tabla de la página adyacente da los 26 grupos simples esporádicos. Las letras que designan las familias infinitas, desde *A* hasta la *G*, proceden de la teoría de Lie, en recuerdo de Sophus Lie. Los nombres con que han sido bautizadas las familias son los de Claude Chevalley, Robert Steinberg, de la Universidad de California, Michio Suzuki, de la Universidad de Illinois, y Rimhak Ree, de Columbia Británica. Los grupos simples esporádicos reciben sus nombres de Emile Mathieu, Zvonimir Janko, hoy en la Universidad de Heidelberg, Donald Higman, de la Universidad de Michigan, Charles Sims, de Rutgers University, Jack McLaughlin, de Michigan, Suzuki y Arunas Rudvalis de la Universidad de Massachusetts en Amherst, Dieter Held, de la Universidad de Mainz, Richard Lyons y Michael O'Nan de Rutgers, John Conway, de Cambridge, Bernd Fischer, de la Universidad de Bielefeld, Koichiro Harada, de Ohio State University, John Thompson, en Cambridge, y Robert Griess, Jr., de Michigan. Los grupos simples individuales que pertenecen a una familia infinita dada se denotan mediante valores numéricos de las variables *n*, *p* y *q*, donde *n* puede ser un entero positivo cualquiera, y *p* un

trabajos de Richard Brauer sobre centralizadores de involuciones establecieron uno de los más importantes hitos en el viaje hacia la frontera. Si bien no todos los elementos de un grupo conmutan necesariamente unos con otros, resulta útil considerar el conjunto de todos los elementos de un grupo que conmutan con un elemento dado a, es decir, el conjunto de todos los elementos g del grupo, para los cuales $g*a = a*g$. Este conjunto se llama centralizador del elemento a. Es fácil verificar que tales centralizadores son grupos cuando la ley de composición de sus elementos es la ley de composición del grupo paterno, y que, por consiguiente, son subgrupos de éste. Un ejemplo trivial es el centralizador del elemento idéntico e, que es siempre el grupo paterno, dado que una de las condiciones de la definición de grupo es que todos los elementos de grupo hayan de conmutar con el elemento identidad de él. Un ejemplo más enjundioso es el centralizador del elemento (123) del grupo de giros del tetraedro: el centralizador de (123) consta de la identidad, del elemento (123) y del elemento (132).

La idea de Brauer consistía en centrarse en aquellos elementos a del grupo, distintos del elemento e, para los cuales el producto $a*a$ es igual a e. Los elementos del grupo que tienen esta propiedad se llaman involuciones, y es fácil demostrar que todo grupo con número par de elementos contiene involuciones. Por consiguiente, en vista del teorema de Feit-Thompson, todo grupo simple no conmutativo contiene involuciones. Brauer comenzó calculando los centralizadores de involuciones en varias de las 18 familias regulares, y observó que tenían la misma estructura general que el grupo paterno, si bien en forma embrionaria. Se preguntó entonces si sería posible reconstruir el grupo paterno exclusivamente a partir del conocimiento de estos centralizadores, y con el tiempo consiguió resolver afirmativamente la cuestión en cierto número de casos importantes.

GRUPOS SIMPLES ESPORADICOS	
NOMBRE DEL GRUPO	NUMERO DE ELEMENTOS
GRUPOS DE MATHIEU	
M_{11}	$2^4 \times 3^2 \times 5 \times 11 = 7.920$
M_{12}	$2^6 \times 3^3 \times 5 \times 11 = 95.040$
M_{22}	$2^7 \times 3^2 \times 5 \times 7 \times 11 = 443.520$
M_{23}	$2^7 \times 3^2 \times 5 \times 7 \times 11 \times 23 = 10.200.960$
M_{24}	$2^{10} \times 3^3 \times 5 \times 7 \times 11 \times 23 = 244.823.040$
GRUPOS DE JANKO	
J_1	$2^3 \times 3 \times 5 \times 7 \times 11 \times 19 = 175.560$
J_2	$2^7 \times 3^3 \times 5^2 \times 7 = 604.800$
J_3	$2^7 \times 3^5 \times 5 \times 17 \times 19 = 50.232.960$
J_4	$2^{21} \times 3^3 \times 5 \times 7 \times 11^3 \times 23 \times 29 \times 31 \times 37 \times 43 \doteq 8{,}68 \times 10^{19}$
GRUPO DE HIGMAN Y SIMS, *HS*	$2^9 \times 3^2 \times 5^3 \times 7 \times 11 = 44.352.000$
GRUPO DE McLAUGHLIN, *Mc*	$2^7 \times 3^6 \times 5^3 \times 7 \times 11 = 898.128.000$
GRUPO ESPORADICO DE SUZUKI, *Suz*	$2^{13} \times 3^7 \times 5^2 \times 7 \times 11 \times 13 \doteq 4{,}48 \times 10^{11}$
GRUPO DE RUDVALIS, *Ru*	$2^{14} \times 3^3 \times 5^3 \times 7 \times 13 \times 29 \doteq 1{,}46 \times 10^{11}$
GRUPO DE HELD, *He*	$2^{10} \times 3^3 \times 5^2 \times 7^3 \times 17 = 4.030.387.200$
GRUPO DE LYONS, *Ly*	$2^8 \times 3^7 \times 5^6 \times 7 \times 11 \times 31 \times 37 \times 67 \doteq 5{,}18 \times 10^{16}$
GRUPO DE O'NAN, *ON*	$2^9 \times 3^4 \times 5 \times 7^3 \times 11 \times 19 \times 31 \doteq 4{,}61 \times 10^{11}$
GRUPOS DE CONWAY	
C_1	$2^{21} \times 3^9 \times 5^4 \times 7^2 \times 11 \times 13 \times 23 \doteq 4{,}16 \times 10^{18}$
C_2	$2^{18} \times 3^6 \times 5^3 \times 7 \times 11 \times 23 \doteq 4{,}23 \times 10^{13}$
C_3	$2^{10} \times 3^7 \times 5^3 \times 7 \times 11 \times 23 \doteq 4{,}96 \times 10^{11}$
GRUPOS DE FISCHER	
F_{22}	$2^{17} \times 3^9 \times 5^2 \times 7 \times 11 \times 13 \doteq 6{,}46 \times 10^{13}$
F_{23}	$2^{18} \times 3^{13} \times 5^2 \times 7 \times 11 \times 13 \times 17 \times 23 \doteq 4{,}09 \times 10^{18}$
F_{24}'	$2^{21} \times 3^{16} \times 5^2 \times 7^3 \times 11 \times 13 \times 17 \times 23 \times 29 \doteq 1{,}26 \times 10^{24}$
GRUPO DE HARADA F_5	$2^{14} \times 3^6 \times 5^6 \times 7 \times 11 \times 19 \doteq 2{,}73 \times 10^{14}$
GRUPO DE THOMPSON F_3	$2^{15} \times 3^{10} \times 5^3 \times 7^2 \times 13 \times 19 \times 31 \doteq 9{,}07 \times 10^{16}$
GRUPO DE FISCHER F_2 ("EL MONSTRUO PEQUEÑO")	$2^{41} \times 3^{13} \times 5^6 \times 7^2 \times 11 \times 13 \times 17 \times 19 \times 23 \times 31 \times 47 \doteq 4{,}15 \times 10^{33}$
GRUPO F_1 DE FISCHER Y GRIESS ("EL MONSTRUO" y "GIGANTE AMABLE")	$2^{46} \times 3^{20} \times 5^9 \times 7^6 \times 11^2 \times 13^3 \times 17 \times 19 \times 23 \times 29 \times 31 \times 41 \times 47 \times 59 \times 71 \doteq 8{,}08 \times 10^{53}$

número primo arbitrario. La variable q tiene que ser igual al número de elementos de un sistema numérico finito, pudiéndose demostrar que para un conjunto de q elementos puede construirse exactamente un sistema numérico finito solamente si q es potencia entera de un número primo. Así pues, los nombres de las familias designan grupos simples tan sólo para valores de q iguales a 2, 2^2, 2^3,... , 3, 3^2, 3^3,... , 5, 5^2, 5^3,... y así sucesivamente, a menos que se observen otras excepciones. Por ejemplo, el grupo lineal de Chevalley $A_1(3)$ queda excluido de la lista por las condiciones impuestas a los valores de n y q, porque $A_1(3)$ es el grupo no simple de giros de un tetraedro regular. El entero n, que se denomina rango de la familia, está restringido a los valores 2, 4, 6, 7 y 8 en los grupos de Chevalley exepcionales. Los grupos de Ree y Suzuki solamente están definidos para valores de q iguales a potencias de 2 ó 3 con exponente m impar. El número de elementos de cada uno de los grupos de la familia está dado tanto por la expresión algebraica como por el número de la segunda columna de cada tabla. Aquí d expresa el MCD (máximo común divisor) de los dos números, o de las expresiones algebraicas que figuran detrás de la abreviatura "MCD". En los ejemplos que se dan en la primera tabla, las barras verticales que encierran el nombre del grupo designan el número de elementos que éste contiene.

El trabajo de Brauer no sólo se encuentra subyacente al descubrimiento de varios de los grupos esporádicos, sino que también proporcionó un procedimiento operativo para subdividir en dos partes el viaje hasta la frontera de la región plana de los grupos simples. Primero, demostrar que el centralizador de una involución en un grupo simple desconocido se asemeja de cerca al centralizador de una involución en uno de los grupos simples conocidos. Los análisis que forman bases para sendas cercanas al punto central están relacionados con este aspecto del problema de clasificación.

Saber que el centralizador de una involución en el grupo desconocido es el mismo que el centralizador de una involución en un grupo conocido no es en modo alguno equivalente a saber que el grupo desconocido sea el mismo que el grupo conocido: el centralizador de una involución no es sino una pequeña porción del grupo total. El propósito de la segunda fase de la demostración estriba en generalizar esta información sobre los centralizadores, que es de carácter local, y convertirla en una equivalencia de carácter global, mostrando así que el grupo simple desconocido es el mismo que uno de los grupos simples conocidos. Los caminos cercanos a los puntos de la frontera tienen por objeto hacer ver de qué forma los grupos simples conocidos quedan determinados por información concerniente a los centralizadores de sus involuciones. La descomposición del teorema de clasificación en estas dos fases puede representarse mediante una curva situada en el interior de la región del plano, aproximadamente a mitad de camino que media entre el punto central y la frontera.

Entre los muchos matemáticos que participaron en el último decenio en trabajos sobre el teorema de clasificación, he de mencionar a Michael Aschbacher, de Caltech (Instituto de Tecnología de California). En 1972, en una serie de conferencias impartidas en la Universidad de Chicago, proponía yo un programa para la determinación de todos los grupos finitos simples que constaba de 16 pasos. Lo que yo hacía, en efecto, era sugerir los caminos que sería preciso seguir en el viaje hasta la frontera. Pronosticaba, además, que el final feliz se produciría hacia finales de siglo; pero mi ampuloso proyecto y el optimismo de mis predicciones fueron recibidos con considerable escepticismo.

Ni los participantes, ni yo, habíamos

tenido en cuenta a Aschbacher. Recién salido del primer ciclo universitario, acababa de iniciarse en este campo y, a partir de ese momento, se convirtió en la fuerza motriz que sostuvo mi programa. Demostró, en rápida sucesión, teoremas sorprendentes, uno tras otro. Aunque en el asalto final tuvieron importante participación otras muchas personas, Aschbacher fue, por sí solo, el responsable de haber reducido la ejecución de mi proyecto de 30 a 10 años escuetos.

¿Qué reserva el futuro? Mi trabajo actual, juntamente con el de varios colaboradores, pretende construir una demostración "de segunda generación" del teorema de clasificación. La primera demostración incluye muchos de los primeros artículos, escritos mucho antes de haberse trazado una estrategia global para rematar la demostración. Así pues, inevitablemente, el desarrollo de la primera demostración contiene algunos arranques en falso, palos al aire y duplicaciones. Si tenemos éxito en nuestro empeño, la longitud de la segunda demostración será solamente una quinta parte de la primera, con la correspondiente y proporcional mejora de claridad conceptual. Según los criterios matemáticos al uso, una demostración de 3000 páginas seguirá siendo una demostración excesivamente larga. Sin embargo, dada la complejidad del problema, una demostración sustancialmente nueva tendrá que esperar a la introducción de métodos enteramente nuevos.

Es probable que la clasificación de los grupos finitos simples tenga amplias ramificaciones matemáticas. El resultado ha sido aplicado ya a campos matemáticos tan diversos como la teoría de algoritmos, la lógica matemática, la geometría y la teoría de números. Se sabe ahora que el grupo monstruo tiene profundas conexiones, aún no del todo esclarecidas, con la teoría de las funciones elípticas.

Menos claro es el impacto que tal clasificación pueda producir fuera del campo de la matemática. Sin embargo, no han faltado especulaciones de que el gigante amable de Griess, el grupo monstruo, participe en la formulación de una posible teoría de unificación de campos, en la física de las partículas elementales. Ahora bien, cualesquiera que sean las aplicaciones últimas, los especialistas en teoría de grupos habrán tenido éxito en resolver la cuestión más vital de su materia, que existía implícitamente desde que Galois añadió a la matemática la noción de grupo.

La alimentación en China

Tras los numerosos altibajos de las últimas décadas, la producción agrícola china comienza a cubrir las necesidades. Que este equilibrio se mantenga depende de la capacidad de la nación para hacer frente a algunos problemas

Vaclav Smil

El problema que se plantea es de resonancia mundial. ¿En qué medida conseguirá China alimentar a más de un quinto de la población del globo con una agricultura que apenas abarca una quinceava parte de las tierras cultivables del planeta? Una situación que se complica por el hecho de que buena parte de las tierras cultivadas están expuestas a oscilaciones climáticas frecuentes y extremas. Mi respuesta al interrogante, basada en largos años de detenido estudio de la agricultura china y sus actividades afines, es que, por el momento, esta nación está saliendo bastante airosamente de la prueba. Que los éxitos logrados consigan mantenerse hasta el siglo XXI dependerá de las respuestas que dé China a diversos problemas, entre ellos el cambio de énfasis en favor de una economía de libre mercado, la erosión de las tierras cultivables y el control de los efectos negativos de la contaminación.

En los largos siglos de gobierno dinástico, la agricultura tradicional china alcanzó admirables niveles de intensidad y complejidad, incluidos elaborados sistemas de regadío y métodos de rotación de cultivos, así como un vasto reciclaje orgánico. Aun así, el sistema tradicional no consiguió asegurar una alimentación adecuada de la población durante las periódicas inundaciones y sequías. El hambre, frecuente, asolaba con efectos devastadores. La situación no mejoró apreciablemente durante los 38 años de la República Nacionalista, que en su mayor parte fueron también años de guerra.

Tras la implantación de la República Popular China, en 1949, el partido en el poder, siguiendo el ejemplo de Stalin, colectivizó la agricultura y comenzó a explotar el campo, en proceso de rehabilitación pero todavía muy pobre, como fuente de financiación de una amplia industrialización. La colectivización se inició en 1953, coincidiendo con la puesta en marcha del Primer Plan Quinquenal. Aunque la colectivización creó perturbaciones sociales que recuerdan la experiencia traumática de la URSS en los años treinta, y a pesar de que las cooperativas agrícolas recibieron unos suministros de producción –abonos y maquinaria moderna, por ejemplo– muy escasos, el primer período de paz interna de que gozaba la nación en el lapso de dos generaciones vino a sumarse a unas condiciones climáticas bastante favorables; todo ello, unido a un gran esfuerzo de los agricultores, permitió un incremento sostenido de las cosechas.

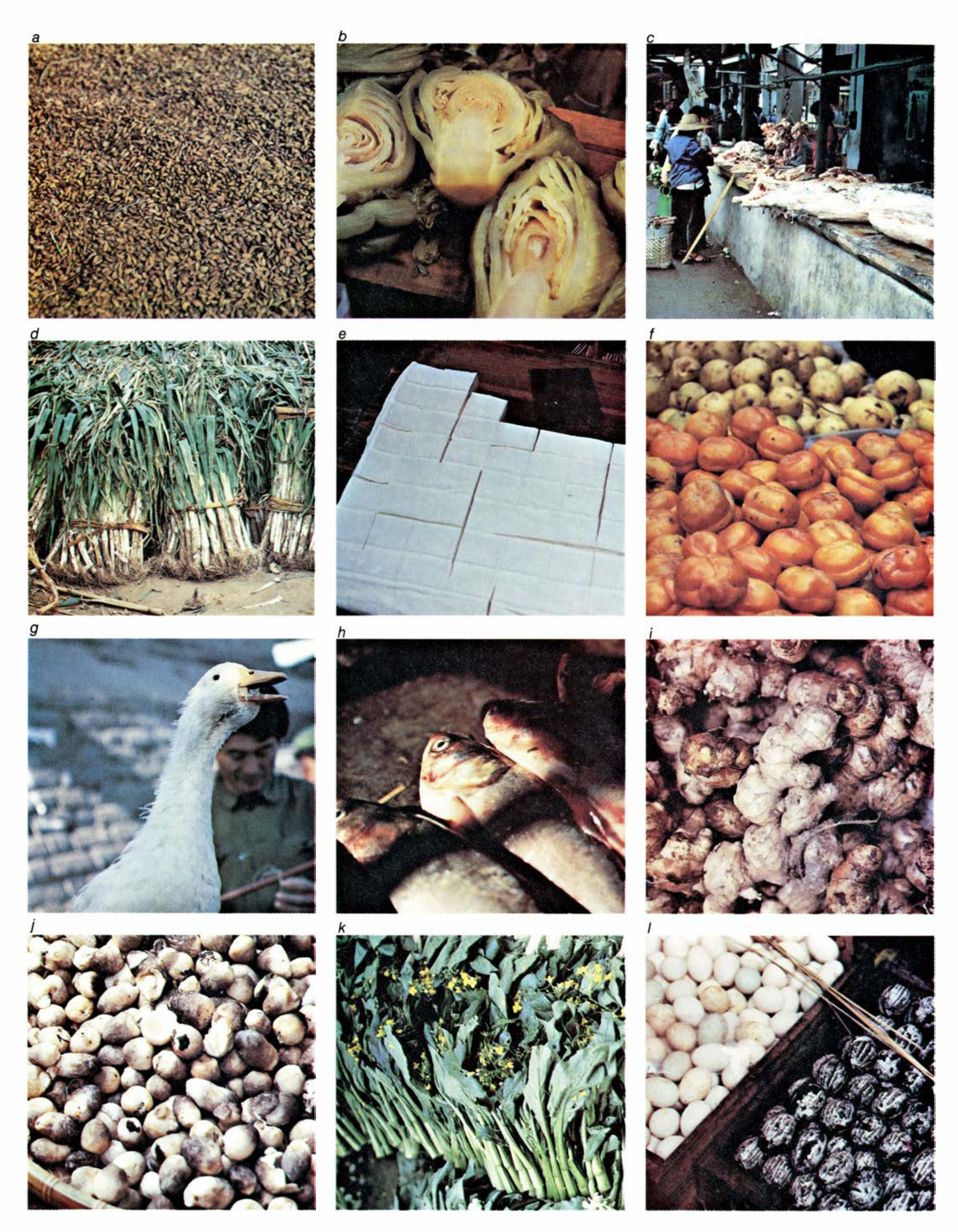
En 1957, la cantidad de alimentos disponibles por persona era superior en un 15 por ciento a los niveles de 1950. Las cosechas de los principales cultivos, excepto la soja, superaron las mejores cosechas totales y los más altos rendimientos logrados antes de 1949, entre 1931 y 1936, concretamente. Parecía muy probable que pudiera seguir manteniéndose un sólido ritmo de crecimiento.

El fracaso del experimento del Gran Salto Adelante de Mao Zedong echó por tierra estas expectativas. Mao se había propuesto alcanzar, en un breve período de tiempo, un nivel de desarrollo que en condiciones normales tardaría varios decenios en llegar. Quiso avanzar a un ritmo demasiado acelerado; impuso presiones exageradas a la economía. En la agricultura, el experimento, agravado por una prolongada sequía en algunas zonas y por extensas inundaciones en otras, se saldó con un retroceso sumamente gravoso para el desarrollo de la nación. Los chinos silenciaron los costes y los observadores extranjeros sólo pudieron aventurar conjeturas; en 1958 nadie disponía de los indicadores económicos y demográficos de rigor. Tendrían que transcurrir más de veinte años antes de que volviera a contarse con estas cifras.

Sólo entonces fue posible la reconstrucción de lo ocurrido, realizada independientemente por John S. Aird, del Ministerio de Comercio de los Estados Unidos, y Ansley J. Coale, de la Universidad de Princeton. Ambos llegaron a la conclusión de que, entre 1959 y 1961, el hambre causó al menos 16,5 millones de muertes adicionales. Los datos demográficos más detallados que se hicieron públicos en China a partir de 1981, en conjunción con los resultados del censo de 1982 (el más fidedigno de los censos chinos, con mucha diferencia), elevan a unos 30 millones el número de muertes adicionales ocurridas en ese período y a unos 33 millones el número de nacimientos frustrados o aplazados. No se tiene noticia de ninguna otra crisis de hambre tan devastadora.

A lo largo de la década de 1960 se produjo una rápida recuperación, seguida, sin embargo, por el desbarajuste de la Revolución Cultural. Durante esos diez años la propaganda oficial reivindicaba reiteradamente grandes éxitos en el campo de la agricultura. Se decía que China había resuelto su problema alimentario. Pero estas pretensiones, al igual que las impresiones recogidas por los extranjeros en fugaces visitas a algunas comunas modelo destinadas a servir de pantalla, no tardaron en quedar desmentidas. A finales de 1977, se publicaron las primeras valoraciones nacionales críticas sobre los resultados obtenidos desde 1957; la primera nueva serie de estadísticas nacionales oficiales se dio a conocer en 1979. A finales de ese año, la nueva actitud de "búsqueda de la verdad basada en los hechos" dio paso a una avalancha de apreciaciones críticas.

Las primeras estadísticas oficiales que se hacían públicas desde 1958 revelaron que, si bien la producción industrial había experimentado un impresionante incremento hasta 1978, la producción de alimentos per cápita no



1. ALIMENTOS BASICOS en la República Popular China, ilustrados en esta secuencia fotográfica en orden aproximado de importancia: (*a*) arroz sin descascarillar, (*b*) coles encurtidas, (*c*) cerdo muy graso, con escasa carne, (*d*) cebollas tiernas del norte, (*e*) cuajada de leche de soja, (*f*) peras y caquis, (*g*) pato de Beijing, (*h*) carpas frescas, (*i*) raíz de jenjibre, (*j*) setas, (*k*) hojas de colza y (*l*) huevos frescos y añejos. El arroz es, con mucha diferencia, el principal componente cereal de la dieta china, el cerdo representa el 92 por ciento del consumo de carne y entre los pescados predomina la carpa. El pato es el ave de corral preferida y las cebollas tiernas y el jenjibre son los principales aderezos de la cocina; las setas y los caquis se sirven frecuentemente como platos especiales.

se había modificado en el plazo de una generación. Nicholas R. Lardy, de la Universidad de Washington, llamó la atención sobre esta notable incongruencia y destacó que China parecía representar el único caso de una nación moderna en la cual la duplicación de la renta nacional real per cápita (equivalente al producto nacional bruto) en un plazo de veinte años aparecía acompañada de un consumo alimentario per cápita constante, o incluso ligeramente decreciente.

La situación ha mejorado apreciablemente desde 1978. Entre esa fecha y 1984, el producto agrícola del país aumentó una media de un 8 por ciento anual. La disponibilidad de cereales básicos per cápita se incrementó en un 30 por ciento, casi se triplicó la de aceites vegetales y casi se duplicó la de azúcar refinado, así como el consumo de carne. Al mismo tiempo, gracias a la multiplicación por 2,5 de la renta real per cápita de las unidades domésticas rurales en el mismo período, los campesinos han podido diversificar más su consumo de alimentos, sin dejar de reducir por ello entre un 10 y un 15 por ciento la fracción de su renta disponible que dedican a gastos de alimentación. Con una valoración conservadora, el contenido energético y proteínico de la alimentación (en términos de una media anual per cápita) se sitúa un 25 por ciento por encima del nivel de 1978.

¿Cómo se ha producido esta extraordinaria transformación, tan rápido avance, sin parangón en cuanto a la combinación de tasas de crecimiento y de objetivos propuestos?



2. OTROS ALIMENTOS CORRIENTES en China son (*a*) los fideos, que aquí se nos ofrecen recién hechos y puestos a secar, (*b*) judías verdes "largas", (*c*) brotes de soja, (*d*) mangos, (*e*) mandarinas, (*f*) pescado seco, (*g*) castañas de agua, (*h*) melón de invierno e (*i*) anís estrellado, popular como condimento.

Para responder a ello citaría dos factores. En primer lugar, un paquete de nuevas medidas reformistas que han ofrecido a los agricultores un mejor abastecimiento de materiales de producción (inputs), unos mercados más amplios y unos precios superiores. También se mitigó la insistencia en el autoabastecimiento local de alimentos. Un énfasis que entorpecía una comercialización regional más eficaz y bloqueaba la especialización local en aquellos cultivos que pudieran explotarse en condiciones ventajosas.

Más importante aún ha sido la política que se resume en el lema de la *baogan daohu*: plena responsabilidad de las unidades domésticas. Un sistema cuyos orígenes se remontan a los años cincuenta y que volvió a reimplantarse con fuerza a partir de 1979, cuando algunas brigadas de producción (las unidades básicas del sistema de comunas rurales establecido en 1958) de las provincias de Sichuan y Anhui contrataron, con las unidades domésticas, el cultivo de cantidades determinadas de productos o la crianza de cierto número de animales. Las unidades domésticas podían conservar el excedente resultante una vez deducidas las cuotas de producción, que se vendían a los precios fijados por el estado, y venderlo igualmente al estado, aunque a un precio más alto, o bien comercializarlo en el mercado libre.

En 1980, esta práctica ya se había extendido a otras provincias; en 1982, la contratación de la producción con las unidades domésticas abarcaba a la mayoría de pueblos. En 1984, la *baogan* se había convertido en la norma habitual para todos los campesinos chinos, salvo un reducido porcentaje. Esta privatización *de facto* de la agricultura china fue el inicio de lo que podría llegar a constituir la transformación socioeconómica de más amplio alcance y más ordenada del siglo XX. Las perspectivas de 800 millones de campesinos pasaron en breve tiempo de un estancamiento de situación de subsistencia a una energica expansión y una incipiente prosperidad.

Se mantiene la propiedad colectiva de la tierra y el estado continúa fijando cupos de producción de los artículos básicos, pero la *baogan* vincula la producción a una recompensa y fomenta la iniciativa, la innovación, la inversión, la eficiencia y la asunción de riesgos por parte de los individuos. Como ha señalado Frederick Crook, del Ministerio de Agricultura de los Estados Unidos, esta práctica concede a las familias cierto grado de control sobre su bienestar y, con él, el orgullo y satisfacción por una buena gestión, el ingrediente crucial de la vida del agricultor que les fue negado durante más de un cuarto de siglo.

A medida que se fue difundiendo por todo el país, el sistema de la *baogan* dejó de limitarse a simples contratos para la producción de determinadas cosechas y cabezas de ganado, hasta incluir todo tipo de actividades rurales, creando así innumerables oportunidades de empleo. Gran número de familias comenzaron a especializarse. Esta tendencia representa una transición altamente significativa de la agricultura de subsistencia a la concentración en las ventajas locales y las capacidades individuales. La mayoría de las unidades domésticas sigue cultivando cereales, pero desde 1982 millones de familias han abandonado la agricultura cerealista en favor de la producción de artículos alimenticios tan especializados como los cacahuetes, las setas y las ostras. También se ha producido una rápida expansión de los servicios especializados. Entre ellos, opciones anteriormente tan inconcebibles como el transporte privado de productos alimenticios a las ciudades y la fumigación aérea de los cultivos. También se observa la aparición de nuevos servicios de información, en forma de asesores locales o itinerantes que aconsejan a los nuevos agricultores emprendedores, dispuestos a asumir un riesgo, sobre formas más rentables de fertilización, reproducción ganadera y gestión de su explotación.

Otro resultado más: por primera vez desde 1953, los campesinos pueden abandonar el campo e incluso se les estimula a hacerlo, aunque la emigración a las grandes ciudades sigue restringida. El incremento de la productividad que acompaña a la *baogan* ha liberado a millones de trabajadores agrícolas, y los están absorbiendo empresas locales de servicios e industrias y se están trasladando a las pequeñas ciudades próximas, donde trabajan en la industria de construcción en expansión, ofrecen una amplia gama de servicios (desde la venta ambulante de artículos para el hogar hasta la fabricación de cuajada de leche de soja y de pasta de sésamo) y abren casas de comidas, casas de té y restaurantes. Comienza a modificarse la relación población rural/población urbana, que se había mantenido estancada en China durante más de una generación. En 1955 ésta era de 87:13 y en 1975 seguía siendo 87:13; en la actualidad se cifra en 75:25.

Continúa ampliándose el alcance de las reformas. Desde 1985, los campesinos ya no están obligados a vender una parte de su cosecha al estado, el cual va renunciando gradualmente al control exclusivo de los productos alimentarios básicos para dejar que la producción se ajuste a la demanda. También se han modificado las condiciones de los contratos para el uso de las tierras cultivables, que solían establecerse para breves períodos. Muchos se han ampliado hasta 15 años de plazo con el propósito de fomentar la introducción de medidas de conservación del suelo y asegurar que se mantenga adecuadamente su fertilidad. En el caso de las explotaciones forestales, los contratos pueden abarcar períodos de entre 30 y 50 años y se garantiza el derecho a la herencia de los árboles. La *baogan*, como ha declarado el viceprimer ministro Wan Li, no es sólo una política práctica para resolver el problema de la alimentación de la población, sino que también representa una reforma fundamental del sistema económico rural.

En estos momentos es posible valorar cuantitativamente con bastante precisión el impacto de esta reforma en el terreno de la nutrición. Aunque en la reciente avalancha de estadísticas oficiales se aprecian algunas discrepancias (y sería ingenuo pensar que se ha eliminado la costumbre de hinchar los datos, práctica omnipresente en las economías centralizadas), las nuevas estadísticas rezuman, sin embargo, una sorprendente sinceridad. Presentan una aceptable coherencia interna, generalmente coinciden con las realidades observables e incluyen gran cantidad de detalles poco halagüeños. Todo lo cual induce al economista a pensar en que, por primera vez en treinta años, dispone de datos adecuados sobre los cuales basar una evaluación realista de los resultados obtenidos en el campo de la agricultura.

Los nuevos datos permiten establecer un balance bastante fidedigno de la producción alimentaria, con descripciones destalladas de la producción, comercialización y variaciones en las existencias de todos los principales alimentos vegetales y animales, teniendo en cuenta sus usos no alimentarios y las pérdidas ocurridas durante el almacenaje y el transporte. Esto nos permite hacernos una idea de la cantidad media

aproximada de alimentos disponibles por familia y por día.

Por mi parte, he realizado estos cálculos para seis años clave de la historia de la República Popular China. En ellos se observa un aumento desde 1800 kilocalorías por persona y por día en 1950 a 2100 en 1957 (al finalizar el Primer Plan Quinquenal), un descenso a 1500 kilocalorías en el momento de mayor azote del hambre en 1960, una recuperación hasta las 2000 en 1965 (el último año "normal" antes de la década de la Revolución Cultural), un pequeño aumento hasta alcanzar las 2100 en 1977 (el último año anterior a la reforma) y un gran incremento hasta las 2700 kilocalorías en 1983.

¿Se puede considerar suficiente esta producción de alimentos? El cálculo de las necesidades en términos de energía alimentaria, para toda una nación, es proclive al error; para que el resultado sea significativo debe expresarse en función de una amplia gama y no de un solo valor. A la hora de calcular las necesidades energéticas me he basado en los datos sobre la distribución por edades, sexo y ocupación de la población china del censo de 1982, cifrando el

MONTAÑAS Y MESETAS DE MAS DE 2000 METROS DE ALTITUD
(33 POR CIENTO)

PASTOS
(33 POR CIENTO)

DESIERTOS
(12 POR CIENTO)

BOSQUES
(12 POR CIENTO)

TIERRAS DE CULTIVO
(10 POR CIENTO)

REGADAS POR LA LLUVIA
(5,5 POR CIENTO)

DE REGADIO (4,5 POR CIENTO)



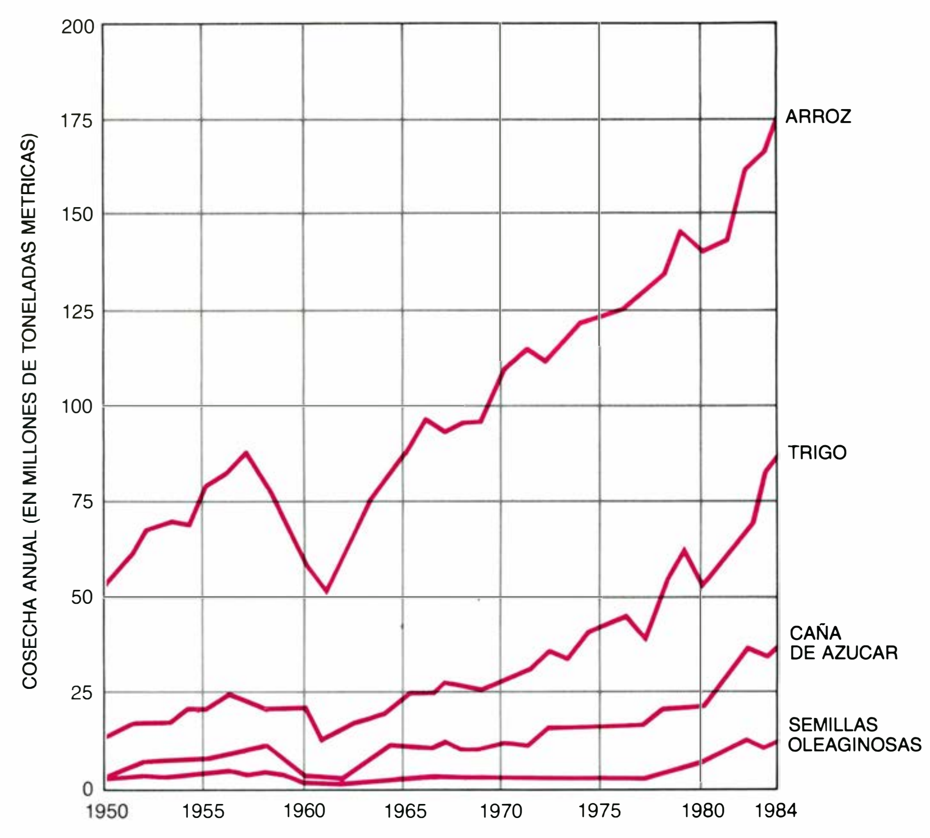
3. EN CHINA, LA AGRICULTURA sólo ocupa alrededor de una décima parte de la superficie del país, debido a las limitaciones impuestas por las condiciones ambientales. Las tierras situadas por encima de los 3000 metros de altitud (y buena parte de las situadas por encima de los 2000 metros) son en gran parte improductivas (*véase el mapa de la derecha*). Una tercera parte del territorio se halla por encima de los 2000 metros. La mayoría de los pastos corresponden a tierras sumamente secas. El mapa también muestra la división administrativa del país y los dos ríos más importantes: el Huang He (Amarillo) y el Chang Jiang (Yangzi o Azul). En el cuadro de la izquierda puede verse la distribución de los distintos tipos de terreno.

peso medio de un adulto en 60 kilos para los hombres y 50 para las mujeres y aplicando las medidas de consumo de energía de los campesinos realizadas por la Academia de Ciencias Médicas china, así como diversas hipótesis sobre la práctica de un gran esfuerzo físico. El resultado son unas necesidades medias diarias per cápita que oscilan entre 2200 y 2400 kilocalorías. (Cálculos de la FAO cifran en 2360 kilocalorías las necesidades medias diarias en China.)

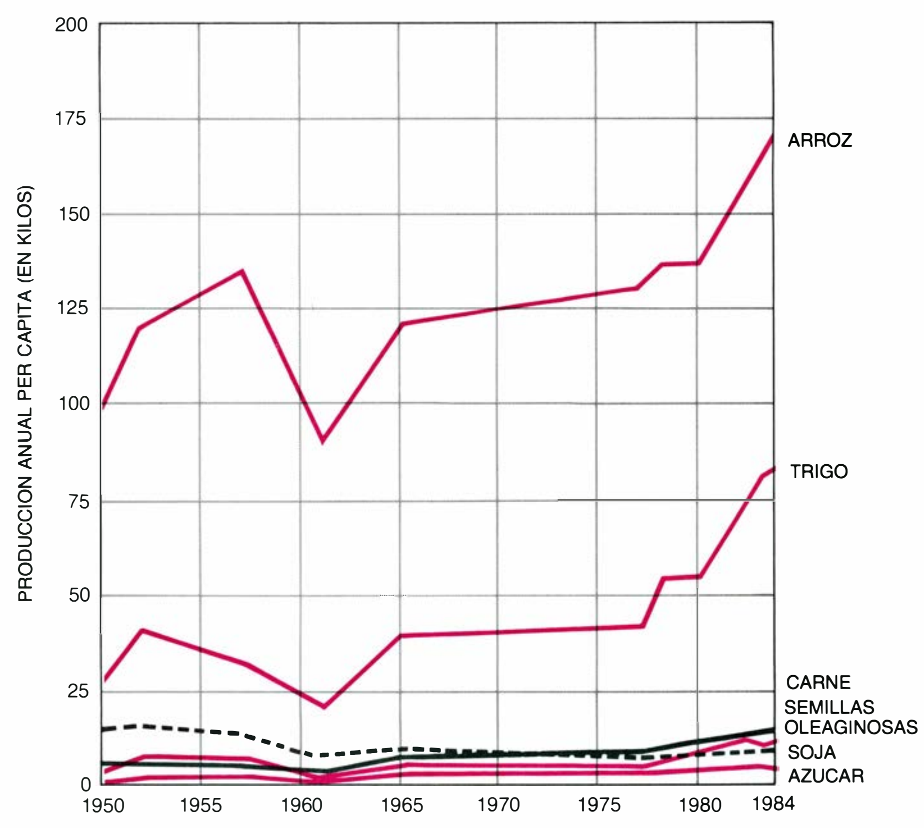
Informes recientes del gobierno de la nación sobre consumo alimentario (el correspondiente a 1983 cubría una muestra de 9060 familias urbanas y 30.427 campesinas) permiten determinar el aprovisionamiento real con una fidelidad sin precedentes. La media nacional ponderada se mueve entre las 2200 y las 2380 kilocalorías per cápita en 1983, es decir, de un 19 a un 12 por ciento por debajo de la oferta media de alimentos.

Lo máximo que puedo deducir de ello –y el lector debe tener presente que este es un campo cuantitativamente incierto incluso en las naciones occidentales– es que actualmente





4. INCREMENTO DE LAS COSECHAS, reflejo de los recientes progresos (un poco desiguales) de la agricultura china. La grave disminución de las cosechas a resultas del Gran Salto Adelante de Mao Zedong provocaron una dura hambrina en 1959-61. Las cosechas de cereales se recuperaron rápidamente, pero otros cultivos de carácter menos básico de la dieta, en particular las semillas oleaginosas, sólo comenzaron a aumentar su rendimiento tras la introducción de reformas en la agricultura, a finales de la década de 1970. Una reforma decisiva es la tendencia actual hacia la implantación de una agricultura de libre empresa.



5. TAMBIEN HA HABIDO UN INCREMENTO en las cosechas per cápita de los principales cultivos. Sin embargo, frente a un incremento de la producción total de alimentos, las cosechas per cápita no llegaron a superar hasta 1977 las cotas máximas alcanzadas en la década de los 50. La mayoría ha mejorado.

se ha logrado un equilibrio básico entre las necesidades y el consumo alimentario medio. Una conclusión que supone un logro significativo. En efecto, desde 1982 China ha producido una cantidad media de alimentos suficiente para cubrir un consumo energético compatible con un crecimiento normal y con una vida sana y activa.

La situación es parecida en lo concerniente a las proteínas. Según mis cálculos –y sin olvidar que, con un consumo de leche prácticamente nulo, con pocos huevos y una cantidad bastante exigua de carne, en su mayor parte grasa, los chinos tienen que obtener la mayoría de las proteínas del reino vegetal, sobre todo del arroz, el trigo y otros cereales–, para cubrir las necesidades de crecimiento y mantenimiento se requiere una media diaria de al menos 60 gramos de proteínas; ésa es la que se consume actualmente en una dieta china media. Cálculos publicados últimamente en China señalan como adecuada la cantidad de 75 gramos diarios. Desde 1982 se ha contado siempre con 75 gramos como mínimo, lo cual nos permite afirmar que las necesidades medias de proteínas están cubiertas.

Pero, evidentemente, el término "medias" encierra el quid de la cuestión cuando se trata de valorar las necesidades energéticas y proteínicas. Sin duda alguna, las disparidades regionales, locales e individuales concernientes al nivel de vida se han incrementado a partir de 1978 y las desigualdades relativas a la nutrición no son nada desdeñables.

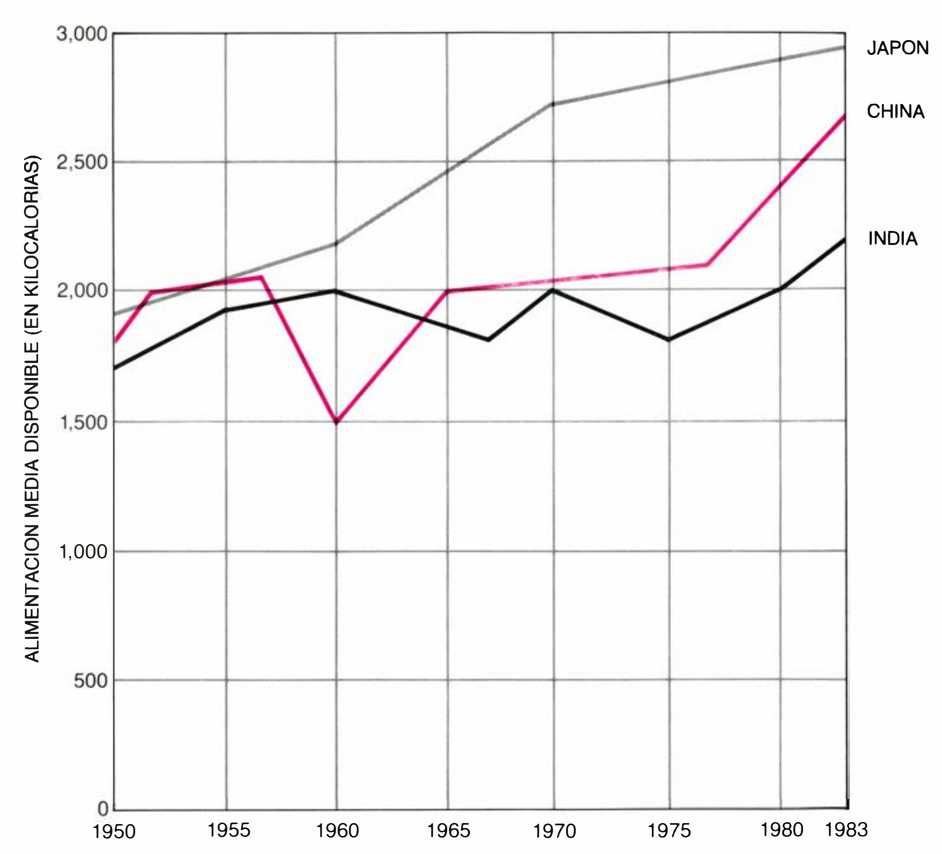
Los dirigentes chinos son conscientes de estas disparidades. En un discurso publicado en enero de 1985, Deng Xiaoping, el principal responsable de las decisiones tomadas, reconocía que "en el campo todavía quedan decenas de millones de campesinos que no tienen suficiente que comer". Un artículo aparecido en *Liaowang* ("Perspectivas") en 1984 recordaba que un "11 por ciento de la población rural todavía no tiene resuelto el problema de contar con ropas de abrigo y una alimentación suficiente". Una cifra que indica que, en 1984, unos 90 millones de personas necesitaban una mejora substancial en su alimentación.

A esto se suman las diferencias, todavía considerables, entre la dieta rural y la urbana. Puede que la población urbana envidie a los campesinos recientemente enriquecidos, sobre todo por sus grandes casas nuevas, pero el ha-

bitante medio de las ciudades todavía sigue comiendo mucho mejor. Igual que en otros lugares, la mejor nutrición de la población urbana en China se caracteriza por un menor consumo de cereales (en 1983, sólo un 75 por ciento del consumo medio rural) y un mayor consumo de grasas (un 85 por ciento de aceites vegetales más que en los pueblos), carne (aproximadamente el doble), huevos (4,5 veces más) y azúcar (más del doble). Pero incluso en las ciudades, la dieta sigue siendo demasiado escasa en proteínas de alta calidad, y más de la mitad de la renta disponible se destina a alimentación (en las ciudades esta proporción es del 60 por ciento).

Otro componente del panorama de la nutrición es un cambio en el tipo de cosechas. Los cereales descascarillados, el arroz de molino y la harina de trigo, que antaño representaban entre el 55 y el 60 por ciento de la producción cerealística total, constituyen hoy un 65 por ciento de la cosecha de cereales. De ahí que incluso las familias rurales de renta media puedan obtener una proporción de productos alimenticios básicos que antes de 1978 sólo formaban parte de las raciones privilegiadas de las ciudades. (Aunque en Occidente estén de moda los cereales integrales y las dietas con un alto contenido en fibras, los chinos manifiestan una marcada preferencia por el arroz bien limpio y la harina blanca; sobre todo después de dos décadas de cereales sin descascarillar y arroz "negro".) En la mayoría de las comidas, los cereales básicos sólo se acompañan de verduras. La variedad de verduras de temporada que pueden obtenerse en los comercios privados ha aumentado considerablemente en los últimos años. Sin embargo, en el norte, durante los meses de invierno, la elección se limita a poco más que las coles, las cebollas y los nabos, y el consumo diario de verduras se reduce a menudo a una tercera parte de la media para el verano.

El avance más favorable es la media de cuatro kilos de aceites comestibles actualmente disponibles por persona y año. Hasta 1978, la escasez de aceite era crónica, con el desplazamiento del cultivo de la colza, los cacahuetes, el girasol y el sésamo en favor de los cereales. Las disponibilidades de aceite disminuían continuamente pese a las importaciones de aceite de soja; en algunas ciudades la ración llegó a situarse por debajo de los 100 gramos mensuales. El actual nivel de consumo urbano, de unos 500 gramos mensuales, a duras penas satisface las exigencias mínimas. (Se requieren al menos 30 gramos de aceite para sofreír sólo un plato de carne o de verduras para una comida familiar media.)



6. TRES NACIONES revelan pautas distintas en la cantidad de alimento per cápita, a partir de posiciones iniciales similares en los años cincuenta. Las cantidades están expresadas en kilocalorías diarias per cápita.

Sólo alrededor de un 6 por ciento de las calorías y un 15 por ciento de las proteínas se obtienen de alimentos de origen animal, cuyo consumo se eleva a unos 25 kilos por persona y año. El cerdo representa más del 90 por ciento de la oferta de carne roja, en su mayor parte en forma de piezas grasas; transcurrirán todavía varios años antes de que llegue a satisfacerse la demanda de piezas magras. Desde 1980, las explotaciones familiares especializadas han venido incrementando la producción de pescado de agua dulce en un 10 a un 15 por ciento anuales. El pescado y el marisco, cuyo consumo supera los tres kilos anuales per cápita, constituyen la segunda fuente de proteína animal en China. El pato al estilo de Beijing (y hasta el simple pollo al vapor) está considerado un manjar exquisito; el consumo medio de aves no llega a los tres kilos al año, ni siquiera entre los 75 millones de habitantes de las principales ciudades. Y aunque el consumo de huevos ha aumentado en un 50 por ciento desde 1978, la media todavía es de menos de un huevo de tamaño grande a la semana por persona.

Por impresionantes que sean todos los avances logrados en la agricultura china, el verdadero problema está en saber si podrán mantenerse o no las tendencias favorables que han venido prevaleciendo desde 1978, no sólo en términos cuantitativos sino también cualitativos. Los obstáculos que se oponen a ello son enormes y la respuesta al interrogante está condicionada a que se consigan mitigar y eliminar, en buena proporción, esos obstáculos.

Un economista occidental no puede dejar de señalar como primera dificultad el carácter irracional del sistema de precios en China. Opinión ésta que comparten los máximos dirigentes, quienes prestan cada vez mayor atención a la necesidad de reformar el sistema de abastecimiento, al mismo tiempo que bajan los precios de los útiles básicos de producción agrícola que se pagan en el campo y suben los precios al detalle de los productos agrícolas en las ciudades. Según la conclusión del análisis de Nicholas Lardy, ni las comunas ni la agricultura *baogan* podrán mantener un crecimiento sostenido de la producción a largo plazo, si continúan topando con precios distorsionados y mercados restringidos.

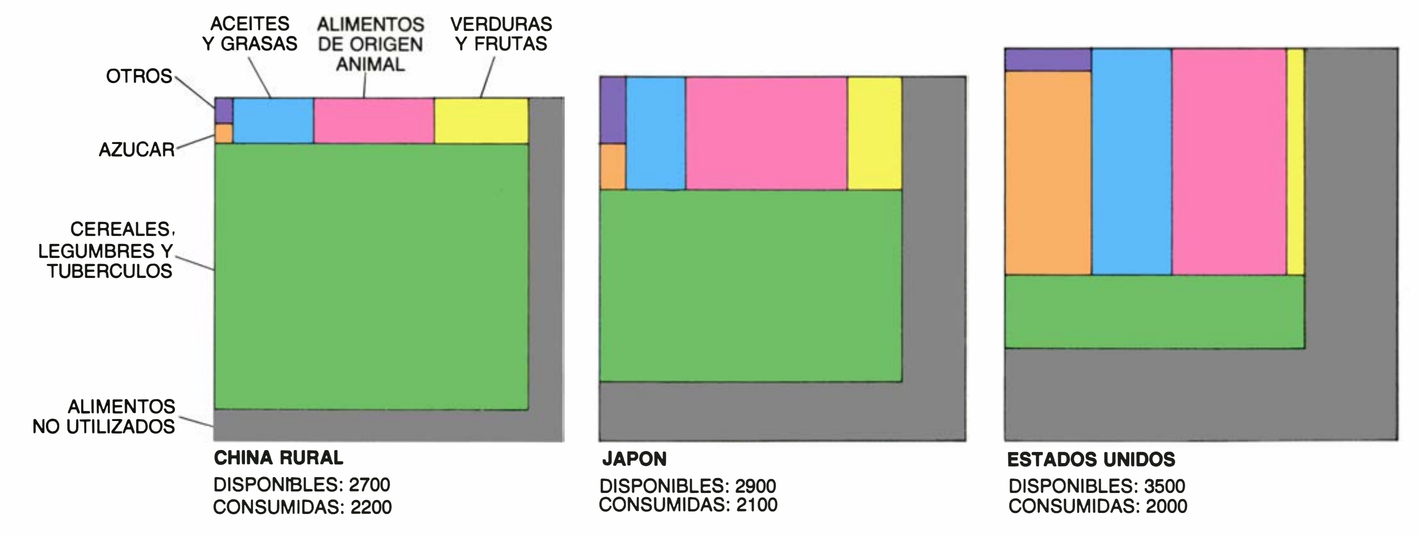
Tal vez el mejor ejemplo en favor de la necesidad de una reforma de los precios fue la reimplantación del racionamiento de la carne de cerdo en las grandes ciudades chinas en febrero de 1985. La baja cotización del porcino convirtió su crianza en una actividad ruinosa y los campesinos la abandonaron en favor de los conejos de angora o los cereales, más rentables. Otro fallo salió a la luz cuando el país no pudo absorber las cosechas récord de cereales de los años 1983 y 1984. Tras años de mero autoabastecimiento local, los agricultores se vieron literalmente en la imposibilidad de deshacerse de su grano, ante una insuficiente capacidad de almacenaje y la falta de un transporte interprovincial eficiente. Durante meses fueron llegando noticias sobre las grandes cantidades de cereales que se estaban pudriendo abandonados a campo abierto. Las deficiencias en el almacenamiento de las cosechas de cereales y una industria de piensos primitiva se traducen en la producción de menos de cinco millones de toneladas anuales de piensos compuestos de la calidad necesaria para una producción rentable de carne magra de cerdo y de pollos. (El producto de la industria de piensos en los Estados Unidos es de casi 200 millones de toneladas anuales.)

El índice general de precios estatales de compra (las cantidades que se pagan a los agricultores) aumentó un 54 por ciento entre 1977 y 1983 con objeto de estimular la producción, pero el índice de precios alimentarios al detalle sólo subió un 15 por ciento. El desajuste ha adquirido tales proporciones que las subvenciones estatales destinadas a frenar el aumento de los precios alimentarios al detalle representan hoy alrededor del 15 por ciento del gasto público total. Las grandes subidas de precios de la primavera de 1985 –un 70 por ciento para la carne magra de cerdo, un 35 por ciento para el cerdo corriente, entre un 100 y un 300 por ciento para algunas clases de pescado– son sólo una parte de los numerosos ajustes de precios necesarios. Más difícil será abordar el problema de unos precios exageradamente bajos de los productos energéticos, una de las principales causas del terriblemente alto consumo de energía en China y un factor de distorsión de los costes de los materiales básicos de producción agrícola.

Incluso con un sistema de precios flexible y una comercialización fluida, la agricultura china continuaría enfrentándose con el enorme desafío de mantener un ecosistema agrícola viable a largo plazo. El anterior énfasis unilateral en el cultivo de cereales y el autoabastecimiento local de alimentos pueden interpretarse como una reacción frente a la hambrina que asoló el país en 1959-61, pero esa política tuvo graves consecuencias. En una nación con un manto forestal ya muy esquilmado, el esfuerzo de "los cereales ante todo" de los años sesenta y setenta dio lugar a una tala indiscriminada. Las provincias más afectadas –Sichuan y Yunnan– perdieron, respectivamente, un 30 y un 40 por ciento de sus bosques en el plazo de una sola generación. La roturación de los pastos del norte para dedicarlos al cultivo de cereales hizo avanzar la desertización. Al rellenar los lagos de las tierras bajas de aluvión, con la misma finalidad, se agravaron las inundaciones provocadas por las riadas estacionales. A partir de 1978, estas desastrosas medidas comenzaron a ser objeto de amplia condena por los científicos chinos, los cuales propugnaban una gestión prudente del ecosistema agrícola. En líneas generales se han dejado de rellenar los lagos y se ha detenido la reconversión de los pastos, pero la deforestación continúa.

Las oportunidades de obtener mayor número de cosechas anuales son limitadas en China. La media nacional del índice de cosechas múltiples, donde 100 representa una cosecha anual por hectárea, ya se sitúa por encima de 150, y oscila entre aproximadamente 100 en el seco y frío noroeste a casi 220 en el curso bajo del Chang Jiang (el Yangzi o Yan-tsé). De hecho, las tres cosechas anuales de cereales que se intentó introducir indiscriminadamente en el este y el sur antes de 1978, sólo pueden mantenerse en contadas partes de estas regiones.

Por todos estos motivos, debería concederse una importancia prioritaria a la conservación de las tierras cultivables. Sin embargo, desde 1957 China ha perdido alrededor del 11 por ciento de sus tierras cultivables a través de la erosión, la desertización y la explotación de actividades no agrícolas. Una extensión que equivale a la totalidad de la superficie cultivable del estado norteamericano de Illinois. La erosión comienza a ser preocupante incluso en la cuenca del Yangzi, antaño relativamente indemne a la misma. La superficie de tierra cultivable per cápita es de 0,1 hectárea, igual a la de Bangladesh y equivalente a sólo dos quintas partes de la media en la India.



7. COMPARACION DE LA DISPONIBILIDAD Y CONSUMO de alimentos de tres poblaciones, en términos de kilocalorías diarias per cápita. La menor diferencia entre cantidades disponibles y consumo se da en China, con una dieta en la que predominan los cereales. Japón y los Estados Unidos presentan pautas notoriamente distintas. Entre los "alimentos no utilizados" se incluyen las cantidades perdidas en el transporte, almacenaje, procesado y desperdicios.

Para mantener unos elevados niveles de rendimiento y proteger sus tierras contra la erosión, los chinos tienen que practicar un intenso reciclaje de la materia orgánica. Así se hacía en la agricultura tradicional china, pero con el aumento de los residuos tóxicos o no biodegradables y un mayor uso de abonos sintéticos se ha ido perdiendo la antigua costumbre, en detrimento de la fertilidad del suelo y de una buena cobertura de mantillo. Resulta particularmente inadecuado el reciclaje de los residuos de las cosechas, esencial para el mantenimiento del suelo, para retener la humedad y como protección contra la erosión. Mientras en los Estados Unidos los agricultores reciclan aproximadamente el 70 por ciento de los residuos de sus cosechas (principalmente paja), en China, con su aguda escasez de energía en las zonas rurales, hasta el 75 por ciento de la paja se emplea como combustible.

Otro preocupante proceso de degradación que amenaza al ecosistema agrícola es la contaminación del agua, causada sobre todo por los residuos urbanos no tratados, por los desechos de las grandes industrias, por los casi cinco millones de pequeñas empresas que existen actualmente en los pueblos y, en un grado creciente, por el uso de abonos nitrogenados. Si consideramos que casi la mitad de las tierras cultivables son de regadío, resulta inmediato ver que las aguas contaminadas por metales pesados, residuos petrolíferos, compuestos fenólicos, hidrocarburos clorados, arsénico y nitratos representan un riesgo muy generalizado de contaminación de los alimentos, así como un peligro directo para quienes las beben. La gran expansión del producto económico prevista para las próximas décadas sólo puede aumentar la contaminación.

China también tiene problemas para satisfacer la demanda de agua de riego en el seco norte. El país dedica 400 millones de metros cúbicos anuales de agua a la irrigación (más del 80 por ciento del consumo total de agua). Se espera que en el año 2000 esta cantidad se habrá incrementado en un tercio y el consumo total de agua en un 50 por ciento. Un nivel de consumo que representaría alrededor de la cuarta parte del caudal total del país. Pero en la mayor parte de las cuencas hidrográficas del norte ya se explotan aproximadamente dos terceras partes del caudal y provincias enteras padecen grandes déficits de agua.

Inevitablemente, China se resentirá muchísimo de estas limitaciones naturales en las décadas venideras, aunque su alcance puede mitigarse bastante con un uso más racional del agua, maximización del cultivo idóneo en la región apropiada, priorización del reciclaje orgánico, implantación de rigurosas medidas de protección de las tierras cultivables y mejora cuantitativa y cualitativa de las infraestructuras y de los aperos de labranza. Pero cada una de estas medidas tiene un coste considerable. Por ejemplo, una irrigación más eficiente exige inversiones masivas en canales revestidos, en aspersores y sistemas de espigas de rotación. Es imposible cultivar las cosechas que mejor se adapten a las condiciones locales a menos que se cuente con un sistema flexible de comercialización y con instalaciones de almacenaje y servicios de transporte a larga distancia de primerísima calidad. Un adecuado reciclaje de los residuos de las cosechas resulta impensable mientras no se alivie la terrible crisis energética en las zonas rurales, principalmente a través de las recién creadas plantaciones privadas de árboles para leña.

Seguirán perdiéndose tierras cultivables de primera calidad, aunque sólo sea para acomodar a una población creciente. Por otra parte, el incremento de las necesidades del campo no resultará barato. Según mis cálculos, los chinos ya emplean en la actualidad un 40 por ciento más de energía por hectárea que los agricultores de los Estados Unidos, a lo cual contribuye en una medida nada desdeñable la baja eficiencia de la conversión de energía.

Dwight H. Perkins, de la Universidad de Harvard, y Shadid Yusuf, del Banco Mundial, plantean el interrogante fundamental al final de un reciente análisis general sobre el desarrollo rural en China: "¿A qué grado de control están dispuestas a renunciar las autoridades y a cambio de qué nivel de incremento en la productividad?" El problema es todavía más profundo. En efecto: ¿se mantendrá la confianza en las políticas reformistas? El propio Deng Xiaoping reconoce que convencer a la población de la irreversibilidad de las nuevas políticas "constituye todo un reto". Sólo cabe confiar que las reformas no sólo se mantengan, sino que se amplíen. Así lo esperamos, tanto por el mejor nivel de vida que ello supondría para la población china, como por la aportación a la estabilidad mundial que representaría este proceso.

Planos del templo de Apolo en Dídyma

El carácter de los "proyectos" con que los griegos construían sus templos se había perdido para los arqueólogos. Un descubrimiento reciente demuestra que estaban trazados sobre la superficie del templo que aquéllos representan

Lothar Haselberger

¿Cómo diseñaban y construían los griegos sus monumentos clásicos que, en el correr del tiempo, se han considerado repetidamente ejemplares? ¿Qué facultades requería la erección de estos imponentes edificios de mármol? ¿En qué teoría estética se fundamentaban? Las respuestas a estas preguntas han sido afanosamente buscadas por los arqueólogos durante más de 200 años. Aunque se sabe que los arquitectos griegos escribían ellos mismos detalladas explicaciones de la construcción y concepción de sus edificios, ninguno de estos escritos ha sobrevivido hasta nuestros días. Ninguna otra fuente revela qué clase de planos de construcción (si los había) eran empleados por los constructores antiguos. A falta de toda descripción de proyectos antiguos, los estudiosos no tenían idea en dónde buscarlos.

Lo paradójico del caso es que los proyectos estuvieron siempre delante de las narices del arqueólogo. Yo mismo he descubierto hace poco todo un archivo de planos todavía *in situ* en un yacimiento que había sido objeto de estudio desde principios de siglo: las famosas ruinas del templo de Apolo en Dídyma, al sur de la actual ciudad turca de Söke.

Durante el viaje por los países de la antigüedad clásica que el Instituto Arqueológico Germánico patrocina anualmente en beneficio de arqueólogos jóvenes, realicé una excursión al templo del Apolo Didimeo. Durante la misma, en octubre de 1979, observé por vez primera multitud de líneas finamente incisas en algunas de las paredes del templo. Lo que me llamó la atención en aquella visita casual a Dídyma se convirtió en objetivo de una empresa arqueológica de alto vuelo, pues es el caso que las líneas en cuestión dibujan una serie de edificios correspondientes al templo. Los diseños incisos tienen una extensión de varios centenares de metros cuadrados y constituyen el conjunto más completo de planos que se conoce de toda la arquitectura antigua.

La construcción del templo de Apolo, el dios griego de la luz, el arte y la profecía, dio comienzo bajo el patrocinio de la poderosa ciudad comercial de Mileto, situada en el extremo más meridional de la antigua Jonia. El templo, sede del oráculo más famoso de los territorios orientales del mundo helénico, estaba destinado a remplazar un edificio anterior (en el sitio de un santuario más antiguo aún) que había sido destruido por los persas.

Los planos del templo empezaron a trazarse poco después del 334 a. de C., cuando Alejandro Magno llegó a Asia Menor. El arquitecto milésico Daphnis colaboró con uno de los principales arquitectos de Jonia, Paionios de Efeso, en el diseño del ambicioso edificio. Terminaba entonces Paionios su obra en el célebre templo de Artemis de la ciudad-estado que lo había visto nacer. Este templo de Efeso, hasta entonces el mayor de los griegos, fue una de las Siete Maravillas del mundo antiguo. El nuevo templo de Apolo estaba previsto que igualase al templo de Artemis en tamaño y esplendor.

La ejecución del proyecto, sin embargo, adelantó muy despacio, y la obra quedó al fin suspendida después de 600 años. En este estado de falta de terminación permaneció durante la Edad Media, hasta que un terremoto derribó la mayor parte de lo que se había edificado.

Aun como ruina el grandioso edificio de mármol me impresionó desde el primer momento. Tres de las columnas se yerguen todavía hasta su altura original de 19,7 metros por encima del basamento escalonado del templo, que mide 120 metros de largo por 60 de ancho. Una de las columnas está sin terminar, bosquejada en su manto de piedra; las otras dos se acabaron y sostienen parte del entablamento marmóreo, la viga en que el tejado estaba destinado a descansar. Un anillo doble de estas altas columnas (108 en total) aspiraba en tiempos a rodear las paredes del recinto interior del templo, todas ellas de la misma altura que dichas columnas. Otras doce columnas labradas, de tamaño y forma similares, ordenadas en tres filas de cuatro, se alzaban en el pórtico, que mira aproximadamente al este. Las diez columnas de la fachada estaban labradas con una riqueza aún mayor que las del pórtico. La mayoría de las columnas estaban ya colocadas en la época en que se suspendió la construcción.

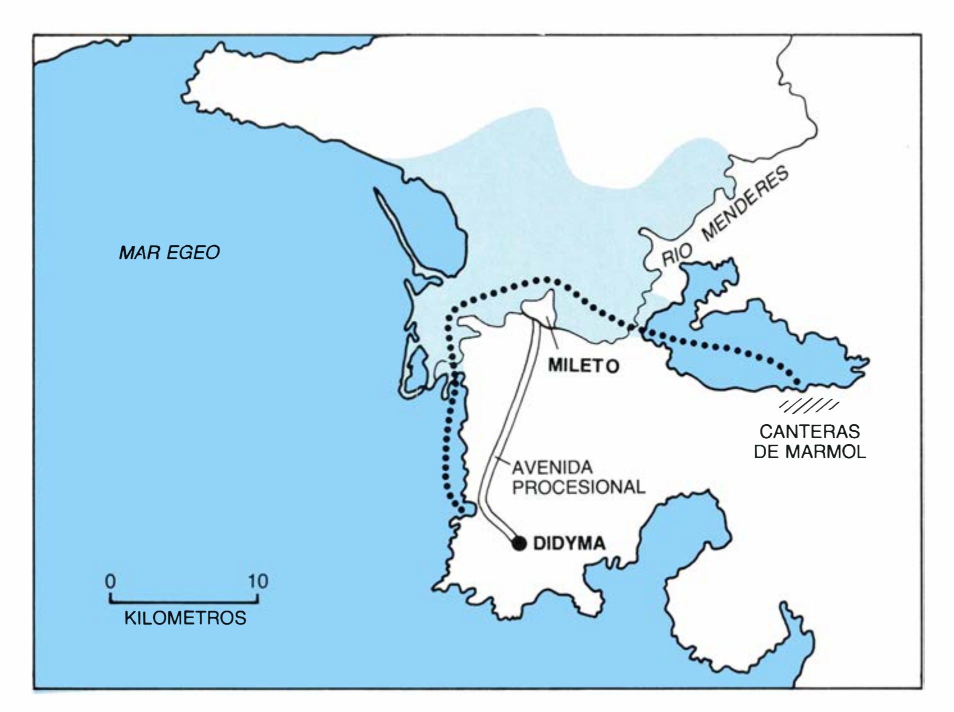
Detrás del bosque de columnas del frontis del templo se elevaba una inmensa puerta, lo que se llama un portal de manifestación. Esta entrada, sin embargo, estaba obstruida: su umbral se alzaba a una altura tal, por encima del suelo, que impedía el acceso. Al otro lado del portal se efectuaban las declaraciones del oráculo desde una sala elevada, a modo de escenario.

Es de suponer que la doble fila de columnas de mármol que orlaba el templo en toda su longitud estuviese destinada a sostener una majestuosa techumbre que cubriese el templo de uno a otro lado. Pero la realidad es que hay un inmenso patio abierto y carente de solería en el interior del templo. Este era el *ádyton*, el área "prohibida", uno de los más notables recintos creados jamás por los arquitectos griegos.

Aquí, a la vera de un laurel, bajo el que se decía que Apolo había sido con-



1. RUINAS DEL TEMPLO DE APOLO en Dídyma, al sur de Söke; impresionan por su propio peso. Una vista frontal (*arriba*) nos muestra las tres columnas subsistentes de las más de ochenta que estaban en pie originariamente. La columna de la izquierda quedó sin terminar: faltan las estrías, o surcos verticales. Las otras dos columnas todavía sostienen una viga de mármol de 5,3 metros de largo que antaño formaba parte del entablamento. El ancho de la base del templo alcanza los 60 metros. El *ádyton* o santuario interior del templo (*abajo*) era un patio grande, sin pavimentar, rehundido en el basamento levantado del templo. Sólo era accesible a través de dos estrechos pasadizos, cuyas salidas se encuentran a ambos lados de las escaleras. Las escaleras daban a una especie de escenario, el "portal de manifestación", desde el cual se emitían los oráculos. Muchos de los trozos de mármol que se hallan diseminados proceden del *naískos*, el templo pequeño del interior. La cimentación rectangular del *naískos* está claramente a la vista (*primer plano, a la izquierda*). También se han descubierto los cimientos de edificios más antiguos. En la parte baja, lisa, de las paredes se hallan grabados planos a escala natural (no perceptibles en estas fotografías) de muchas de las edificaciones del templo. La parte superior de las paredes ha sido reconstruida en parte hasta aproximadamente un tercio de su altura original.





2. DIDYMA se encuentra en la parte más meridional de la antigua Jonia. El mapa pequeño (*a la izquierda*) sitúa el área ampliada (*arriba*) en relación a la Grecia y la Turquía modernas. La ciudad tutelar de Dídyma, Mileto, estaba unida a ella por una avenida procesional de 20 kilómetros de largo. Los bloques de mármol del templo fueron extraídos de una cantera de las montañas del sudeste de Mileto y transportados por mar un trecho considerable (*línea costera de puntos*). Mileto ya no es un puerto; la región sombreada de azul claro es ahora tierra.

cebido, brotaba del suelo un sagrado manantial oracular. Sobre la fuente se alzaba un templete interior primorosamente construido, el *naískos*, donde se custodiaba una estatua de bronce del dios. Aunque el templete fue arrasado a comienzos de la era cristiana, ha sido posible reconstruirlo en el papel a partir de fragmentos.

Para entrar en el *ádyton* (cosa que sólo a unos pocos elegidos se les consentía en la antigüedad) es necesario recorrer uno de los dos pasadizos estrechos, obscuros y en cuesta, situados a ambos lados del portal de manifestación. Cada uno de ellos lleva a una antecámara parecida a una tumba, y cuando se sale de una de estas cámaras al *ádyton*, se queda de momento ciego por la luz repentina.

Las paredes del *podium* que rodea la base del *ádyton* bastan para dejar fascinado al visitante. A decir verdad, lo que a mí me intrigó al principio fue la impecable precisión con que habían sido ajustados los bloques de piedra de las paredes. También aquí, en la pared y a sólo unos pasos de la salida del pasadizo del lado norte, fue donde me llamó la atención una multitud de líneas finamente incisas. Me chocaron en particular los estrechos intervalos regulares, de 1,8 a 1,9 centímetros, intercalados entre ellas.

El resplandor de la luz del sol reflejado por la blanca pared de mármol me deslumbraba y el primer día no pude examinar las líneas con detenimiento. Volví al día siguiente, cuando la luz incidía sobre la superficie con un ángulo más favorable. Esta vez se dejaban ver más líneas, y también semicírculos y cuartos de círculo. El descubrimiento me obligó a pasar más días escrutando las paredes del *ádyton* y de otros lugares de las ruinas.

Descubrí que una delicada red de líneas superpuestas y de composiciones geométricas cubre la mayoría de las paredes del *podium* que rodea el *ádyton*. Otras paredes lisas del monumento, que originalmente estaban al alcance de la mano, son portadoras también de líneas semejantes. Entre unas y otras los dibujos cubren un área de unos 200 metros cuadrados.

Hay líneas rectas de hasta 20 metros de longitud y círculos cuyos radios alcanzan hasta los 4,5 metros. Se han trazado líneas paralelas, polígonos y ángulos subtendidos, y las distancias han sido subdivididas con exactitud. Las líneas son tan delgadas como trazos de lápiz, incisas en la superficie del mármol poco más de medio milímetro. Las rectas y las curvas se trazaron minuciosamente con un fino punzón metálico y con ayuda de una regla larga, en el primer caso, y de compases, en el segundo. También pueden verse líneas imprecisas, pero éstas se corrigieron siempre.

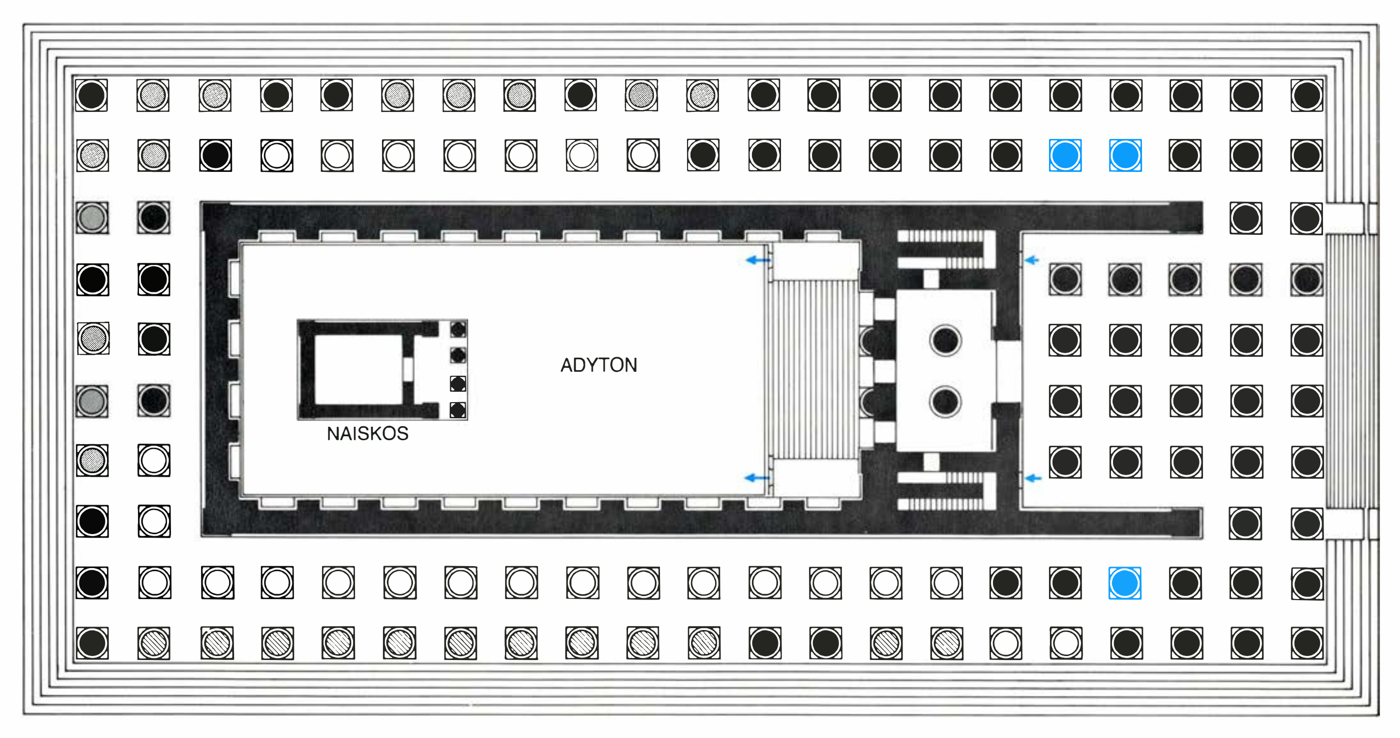
Tanto las líneas como los lienzos murales en que están grabadas se han visto muy poco afectados por la exposición a la intemperie. Esto puede sorprender. En algunos sitios han sido erosionadas por el agua de la lluvia y en otros cubiertos por la toba, la delgada capa mineral que deja el agua al evaporarse. En algunos casos la superficie de la piedra se ha desprendido y las líneas se han perdido. Sin embargo, en otros muchos sitios, las líneas aparecen como si las acabasen de trazar. A pesar de lo cual, incluso donde existen cantidades de líneas legibles, la figura que forman, como un todo coherente, no se puede percibir con la vista: en cuanto uno se aleja un poco, para obtener un campo de visión mayor, las líneas se hacen rápidamente indistinguibles.

Los numerosos residuos de pigmento mineral rojizo que se han conservado en estas paredes (principalmente debajo de sedimentos de toba) indican que la superficie a dibujar se coloreó originariamente de tiza roja, para que resaltasen los diagramas. La tiza debió de ser aplicada antes de que los dibujos se hiciesen, de modo que las líneas incisas se verían blancas sobre un fondo rojo obscuro, produciendo así perfiles de precisión muy nítida. Además de eso, con sólo aplicar más tiza era fácil "borrar" y corregir.

Desde el principio me di cuenta de que estos bocetos no se podían desechar como rayaduras carentes de sentido. Su calidad y cantidad obligaban a considerarlos obra de expertos artífices.

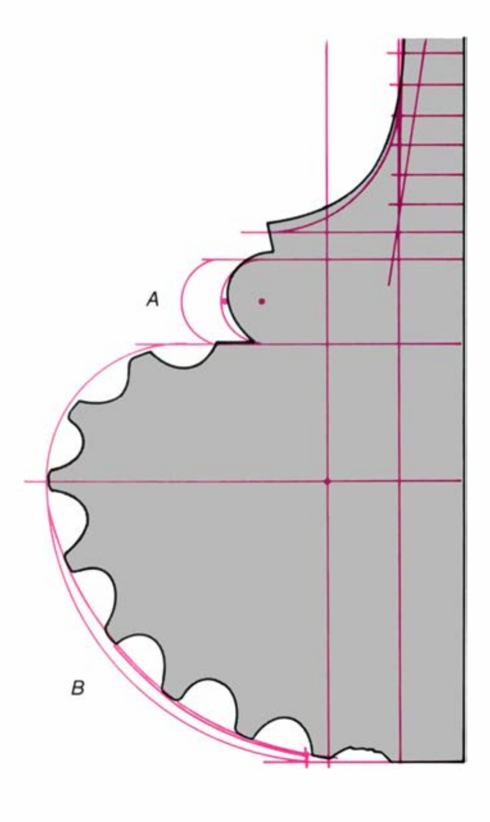
Hoy estas líneas no se pueden ver

3. DIBUJO DEL TEMPLO, con el aspecto que pudo presentar antes de que su construcción fuese interrumpida. Realizó el dibujo George Niemann en 1912, mientras el templo se estaba aún excavando. La reconstrucción adolece de inexactitudes: las columnas están dibujadas como cilindros con una ligera disminución, pero rectangulares; e igualmente, el basamento escalonado, así como las vigas, están dibujados como elementos rectilíneos. En realidad, el templo entero, de los cimientos al tejado, se "abombaba" ligeramente hacia arriba y las columnas se "hinchaban" levemente hacia afuera. Además, las columnas (compuestas de segmentos en forma de tambor, apilados unos sobre otros) fueron erigidas antes de que se esculpiesen y puliesen las 24 estrías que las recorrían.



4. PLANO HORIZONTAL DEL TEMPLO: representa las 122 columnas (*círculos*) proyectadas para él. Las columnas pintadas de negro fueron levantadas en todo o en parte; lo mismo puede decirse con probabilidad de las columnas grises. Las columnas rayadas estaban en fase de erección también cuando la construcción se paró, aunque esto puede afirmarse con menos seguridad. Las columnas dejadas vacías no habían sido aún levantadas. Las tres columnas que están en pie tienen color. El *naískos* se alzaba en el *ádyton*, carente de tejado y de solería. Los dos pasadizos que llevan al *ádyton* están señalados por flechas.

dentro menos de un centímetro la parte inferior del toro para hacer compatible la figura con la verdadera sección transversal del toro.





5. DIBUJO Y BASA REAL DE LA COLUMNA comparados en sus perfiles. El perfil en gris (*boceto de la izquierda*) es diagrama de la sección transversal, a escala, de la basa de una columna del templo (*fotografía de la derecha*). Las líneas incisas en la pared del *ádyton* se hallan aquí dibujadas (*color*) a la misma escala que la sección, hasta donde son visibles o pueden reconstruirse con confianza prolongando las líneas visibles. Despreciando las que parecen ser "primeras pruebas" en el dibujo de una basa, la comparación resulta ajustada. La repetición del dibujo de la moldura semicircular (*A*) y de la curva saliente por debajo de ella (*B*) indica que los bocetos de la pared eran proyectos cuyas versiones definitivas se ejecutaban de inmediato.

con nitidez a menos que la luz del sol proyecte sombras en su interior; lo que se puede hacer es esperar la hora apropiada del día o reflejar luz con espejos. Dado que ningún procedimiento fotográfico permite lograr la exactitud del detalle, la única forma de conseguir una imagen de conjunto de estas líneas ha sido copiarlas a escala después de medirlas cuidadosamente.

La tarea de examinar, medir y registrar las líneas fue iniciada en 1980 como parte de un estudio de conjunto de Dídyma dirigido por Klaus Tuchelt y patrocinado por Edmund Buchner, presidente del Instituto Arqueológico Alemán. En el espacio de 39 semanas de trabajo meticuloso y paciente pude pasar al papel aproximadamente la mitad de las figuras.

Una vez transcrito el primer dibujo completo dimos rápidamente con la solución del primer problema planteado por el descubrimiento: ¿Qué representan los dibujos? Si se compara el contorno o perfil de la basa de una columna con lo que se copió en una de las paredes del *ádyton* la semejanza resulta obvia inmediatamente [*véase la figura 5*]. Este fue de hecho el fulcro de Arquímedes que nos puso sobre la pista correcta, indicándonos que los dibujos son planos del templo y de su *naískos*. Las partes individuales del edificio fueron trazadas en las superficies de las paredes a tamaño natural (a escala de 1/1) y con la máxima precisión.

Aunque sea sorprendente la semejanza entre el bosquejo de la basa de la columna y la sección transversal real de la misma, no deja de haber diferencias entre aquél y ésta. La más llamativa es a la vez la más fácil de explicar: los surcos horizontales de la parte bulbosa y saliente de la basa, llamada toro, no aparecen en los dibujos de la pared. Los surcos se esculpían siempre después de dar forma a los contornos de la basa de la columna. Sólo así se podían hacer correctamente.

Ciertos rasgos menos llamativos son más interesantes. La moldura semicircular situada encima del toro se dibujaba dos veces, primero en una posición proyectada hacia delante y después retrasada 2,4 centímetros, en su verdadera posición. Además de esto, dos líneas curvadas desplazan hacia dentro menos de un centímetro la parte inferior del toro para hacer compatible la figura con la verdadera sección transversal del toro.

Estos cambios menudos pueden parecer triviales, especialmente considerando las enormes dimensiones de la columna (cerca de 20 metros de altura y dos metros de espesor). Sin embargo, tales reajustes no son meras bagatelas en el contexto de los rígidos criterios de forma y proporción a que los arquitectos griegos se creía eran adeptos. La plantilla para diseñar un componente cualquier de un templo tenía que obedecer a muy estrictas y precisas relaciones geométricas.

El dibujo del perfil de la basa de la columna, por ejemplo, estaba determinado por ejes de coordenadas rectangulares [*véase la figura 6*]. El eje vertical estaba destinado a sólo tocar la superficie del fuste de la columna, y el eje horizontal especificaba dónde el toro alcanzaría, a lo largo del eje vertical, el punto de proyección radial más distante. Las otras dos líneas horizontales señalaban el tope y el fondo del perfil del toro; la más baja de ellas estaba exactamente a una distancia del eje horizontal que era el doble de la distancia de la más alta.

Lo que seguía en la construcción del perfil "perfecto" del toro responde a simples reglas y relaciones geométricas. Una curva ajustada al borde externo, redondeado, del perfil del toro puede construirse fácilmente inscribiendo un círculo pequeño en otro mayor, de tal manera que el punto en que sean tangentes, así como sus centros, descanse en el eje horizontal. Si trazamos un cuarto del círculo pequeño dibujando hacia arriba desde el punto de tangencia y después trazamos un cuarto del círculo mayor dibujando hacia abajo desde el punto de tangencia, la curva así dibujada se amolda a la curva del toro en su perfil. Tal y como está dibujado en la pared de Dídyma, el círculo mayor está centrado en el origen de los ejes de coordenadas (el agujero para el compás es todavía visible) y tiene un radio igual a la distancia entre el eje horizontal y la línea horizontal trazada por debajo de él (dos tercios de la altura del toro). El círculo pequeño tiene un radio igual a la distancia entre el eje horizontal y la línea trazada por encima del mismo (un tercio de la altura del toro). Una línea vertical, dibujada hacia arriba desde el centro del círculo más pequeño, debería ser tangente al borde exterior de la moldura

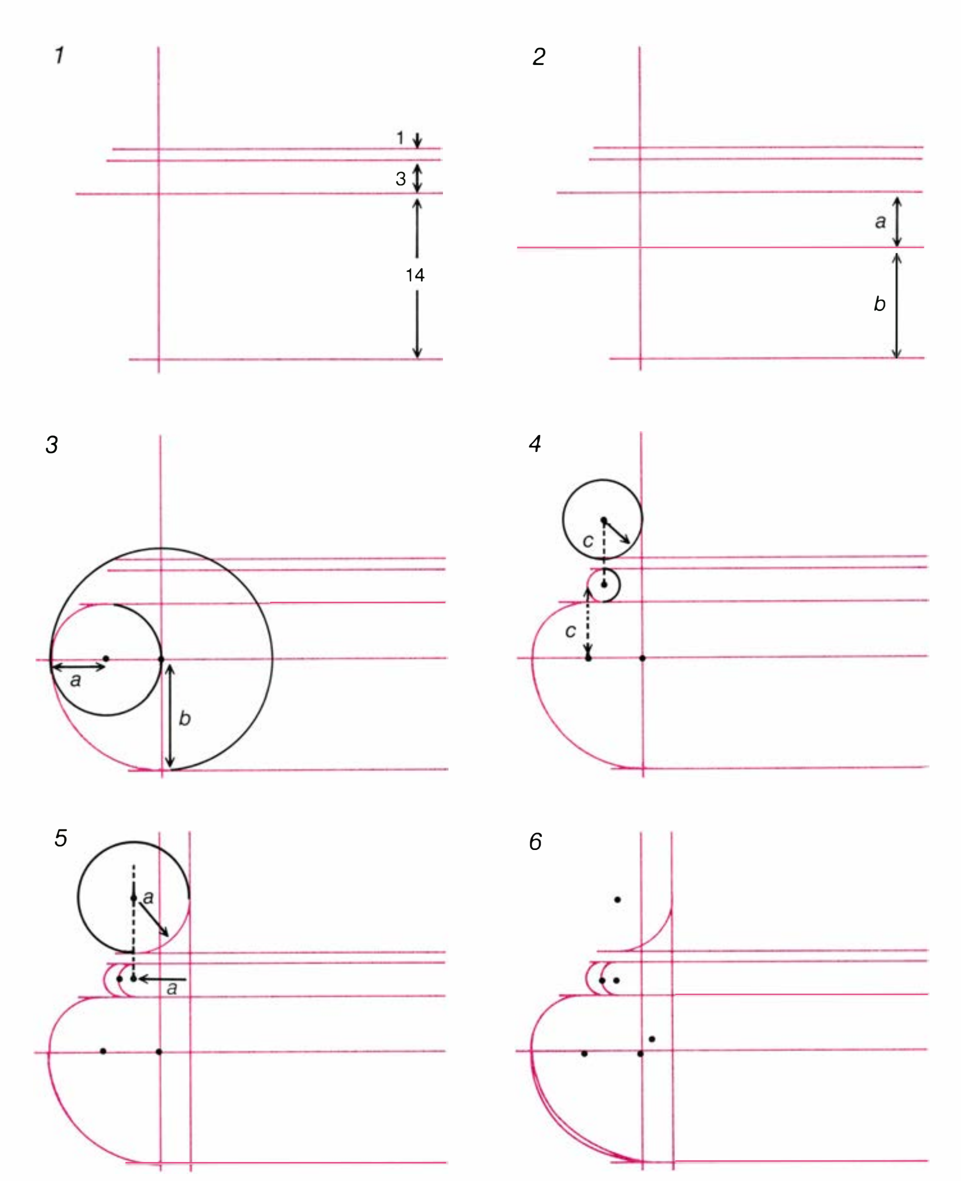
semicircular de encima del toro. Una línea vertical, desde el centro de la moldura semicircular, debería determinar, a su vez, la curvatura y el punto de tangencia del surco hueco que, desde él, se remonta hacia arriba hasta el fuste de la columna.

Este diseño geométricamente perfecto no era, sin embargo, el definitivo, sino que, además, se refinaba de un modo incompatible con la estricta plantilla previa. La moldura semicircular se retraía ligeramente, de su posición fijada geométricamente, para producir un efecto menos apiñado, más nítido. La parte inferior del toro se remetía, con ayuda de trazos sueltos de compás, de radio y posición indeterminados, dando por resultado una curva más tirante.

Aparentemente, el maestro constructor, responsable del diseño (no cabe dudar de que era maestro en su arte), se dejaba guiar, pero no atar, por las estrictas obligaciones que le imponía la composición geométrica. No obstante, estas reglas que se imponía a sí mismo, se las saltaba cuando la estética lo exigía. Esto sin rechazar del todo las proporciones fundamentales del diseño griego. Una vez que la moldura semicircular era retraída, el eje de coordenadas vertical (que representaba el fuste de la columna) tenía que volverse a dibujar. Y de hecho, el nuevo eje vertical era dibujado, otra vez, a una distancia de un tercio de la altura del perfil del toro (la misma unidad que figuraba de modo tan prominente en el proceso original del proyecto), desde el centro de la moldura semicircular desplazada.

Las recíprocas consideraciones de reglas estrictas y arbitrario capricho obligaron al maestro constructor a realizar un balanceo arquitectónico entre dos contrarios. Un áureo término medio entre dos extremos es precisamente lo que los filósofos griegos recomendaban como fuente fundamental de la belleza en su más amplio sentido. Evidentemente, también la arquitectura estaba penetrada, hasta sus menores detalles, por una lucha para alcanzar la belleza del mismo modo.

Es claro que estos "dibujos de trabajo" se empleaban para realizar las partes que componían el templo, en un proceso de diseño muchas veces sumamente complicado. Primero, se dibujaban paradigmas "puros", desde el punto de vista geométrico; pero los arquitectos no vacilaban en alterarlos de acuerdo con la refinada sensibilidad de cada uno. A partir de entonces, la obra de construcción adelantaba de acuerdo con los resultados. La secuencia visible de diseño, revisión y realización indica que, desde luego, los dibujos no se grababan en las paredes como ilustración de lo que ya se había construido. Con ello es posible establecer la cronología de estos diseños con mayor precisión. Los planos de la construcción hubieron de ser dibujados antes de que los distintos componentes se hubieran ejecutado, pero después de que se alisaran los "tableros de dibujo" de piedra.



6. PROCESO DE DISEÑO DE LA BASA DE UNA COLUMNA. Constituía una combinación de rígida construcción geométrica y de refinamiento subjetivo. Cuatro líneas horizontales se dibujaban a través de una línea vertical (*1*), de modo que las distancias respectivas entre cada par de líneas fuesen (*de arriba abajo*) un dáctilo (1,85 centímetros), tres dáctilos y 14 dáctilos. Se trazaba otra línea horizontal (*2*), que partía la distancia entre el par de líneas del fondo, de tal modo que *a*:*b* fuese igual a 1:2. En torno al punto de intersección de esta línea con la vertical, se trazaba un círculo de radio *b* (*3*). Entonces se inscribía en el círculo mayor un círculo de radio *a*, más pequeño, de modo que su centro y el punto común de tangencia estuviesen también en la línea horizontal. A partir de este punto de tangencia, y dibujando hacia arriba, se trazaba un cuarto del círculo menor; arrancando de nuevo del mismo punto de tangencia, pero dibujando hacia abajo, se trazaba un cuarto del círculo mayor. La porción dibujada pretendía representar la sección transversal de una basa de columna geométricamente "perfecta", llamada el toro (del latín *torus*, cojín). Una línea vertical trazada desde el centro del círculo menor (*4*) estaba destinada a fijar la posición de la moldura semicircular dibujada encima del toro, entre la pareja central de las líneas horizontales originales. El centro de la moldura, a su vez, se utilizaba como punto terminal de otra línea vertical igual a *c*. Un cuarto de círculo centrado en el otro punto terminal estaba destinado a definir la curva ascendente que conectaba la basa de la columna con el fuste de la misma. Sin embargo, antes de continuar con este dibujo, se produjo un cambio arbitrario (*5*). La moldura semicircular se retrasó una distancia corta, que al parecer no guardaba relación con ninguna de las proporciones geométricas que se habían empleado hasta aquel momento. Sin embargo, el dibujante volvió a las proporciones originales cuando retrasó la línea vertical hasta una distancia *a* del nuevo centro de la moldura. De este modo, la nueva curva que conecta la basa de la columna con el fuste puede definirse mediante un círculo de radio *a*. El bosquejo definitivo, tal como está en la pared que yo descubrí (*6*), incluye también varias curvas trazadas a mano alzada, que recubren en la parte baja dos tercios del toro.

Gracias a una de las "memorias anuales" que han sobrevivido de esta gran empresa (las autoridades del templo exponían al público losas de piedra en las que se resumían los progresos del edificio y los costos hasta la fecha), sabemos que las primeras columnas se erigieron hacia mediados del siglo III a.C. El dibujo de las columnas y del perfil de sus basas hubo de preceder, algún tiempo breve por lo menos, a aquel acontecimiento. Según otras relaciones de la edificación, las paredes del podio del *ádyton*, donde fueron dibujados los modelos de la columna acabada y de la basa, se construyeron más o menos en la misma época, hacia el 250 a.C. Por ello, podemos datar con confianza, en este período, estos y todos los dibujos afines.

Fue por entonces, ochenta años después de que el suelo para el templo fuese roturado por primera vez y de que se terminase en su mayor parte la obra de construcción del basamento escalonado, cuando se pudo prestar atención al diseño detallado de las columnas y de las paredes de la fábrica. Y, en efecto, dado que las paredes del *ádyton* fueron las primeras en ser erigidas, la mayoría de los dibujos se encuentran en ellas. Para entonces, los arquitectos autores del proyecto, Paionios y Daphnis, llevaban mucho tiempo muertos.

Siendo las columnas el rasgo dominante de los templos griegos, no se escatimaban los medios para su fabricación. Para el templo de Apolo hacían falta no menos de 120 columnas. Por los registros de contabilidad sabemos que cada columna costó alrededor de 40.000 dracmas o, a los costos de hoy, cerca de un millón de dólares. Gran esfuerzo había que dedicar también al diseño de la columna.

Para que fuese posible hacer un plano de detalle, se dibujaba a tamaño natural, en las paredes del *podium* del *ádyton*, el fuste de estas columnas, con sus casi 18 metros de largo. Naturalmente, como las paredes no miden 18 metros de altura, había que dibujar la columna acostada.

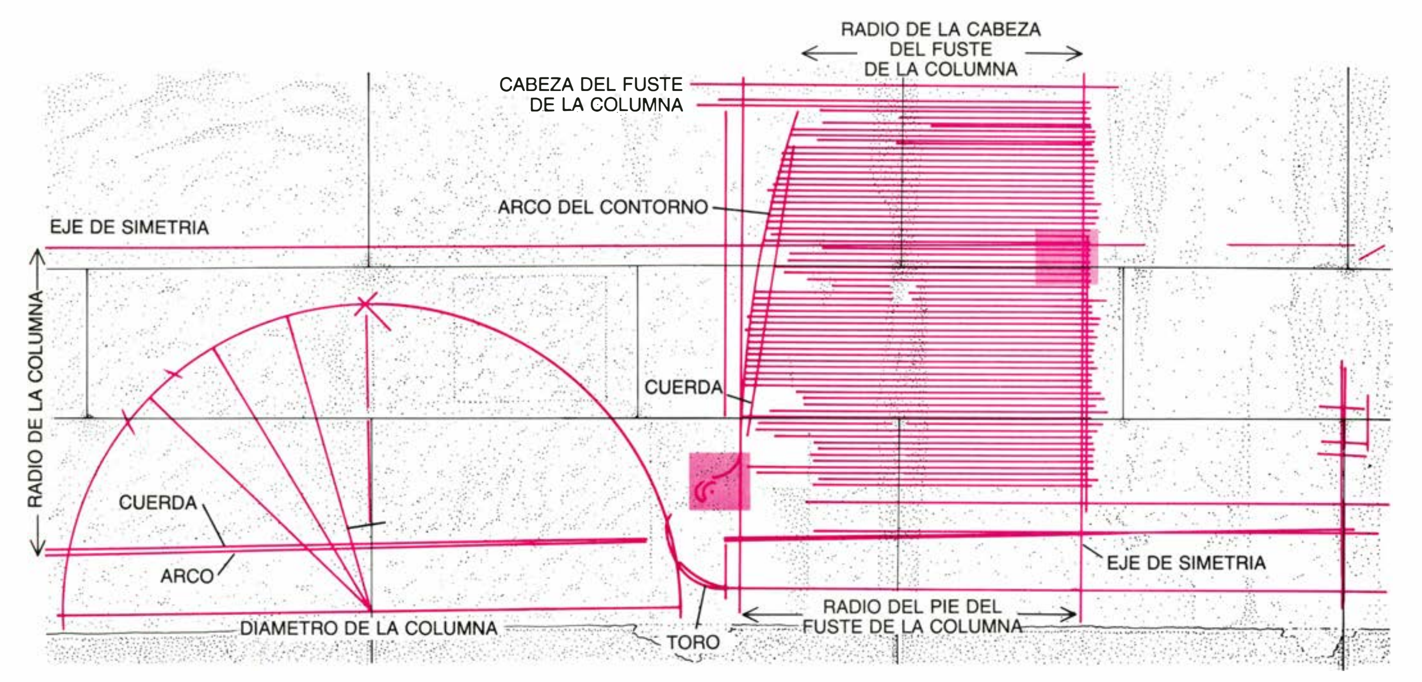
Tres líneas bastaban para representar la forma escueta del fuste de la columna. Un par de líneas muy próximas entre sí, una recta y la otra ligeramente curvada, divergen una de otra y después de alcanzar una separación máxima de 4,65 centímetros en la mitad de su longitud, vuelven a reunirse. Estas serán las líneas de construcción de la curvatura casi imperceptible del fuste. La tercera línea se encuentra a cierta distancia por encima de las otras dos, y discurre casi paralela a ellas. (La distancia entre esta línea y las líneas emparejadas difiere sólo un sexto del uno al otro extremo de la columna.) Dicha tercera línea representa el eje central de la columna; también es el eje de simetría de la columna. Aunque estas líneas constituyen sólo la mitad de una sección transversal vertical, son suficientes, porque la otra mitad no es más que su reflejo simétrico. Los artífices que grababan estos dibujos sólo representaban, en casi todos los casos, una de dos mitades idénticas.

La sección transversal, vertical y diagramática, de la columna se complementa con una sección transversal horizontal que también se bosquejaba de forma simplificada junto con sus ejes de simetría. Un semicírculo grande, de las dimensiones apropiadas, con tres divisiones radiales espaciadas igual, bastaba para especificar la distribución de las veinticuatro estrías de la columna, o surcos verticales.





7. LINEAS INSCRITAS, que representan particularidades constructivas del templo de Apolo, pueden verse en este par de fotografías. La curva (*en la de la izquierda*) que desciende hacia un par de semicírculos, define el pie del fuste de la columna al juntarse con la basa de la misma sobre una moldura semicircular. El semicírculo de la derecha corresponde a la posición real de la moldura en la basa. Encima, y a la derecha, de estas curvas grabadas, hay una serie de líneas paralelas (*fotografía de la derecha*). Son divisiones de la unidad de medida, cada una de ellas representando un pie griego, de longitud real (29,6 centímetros), a lo largo de un boceto a escala de una columna. Se aprecian con claridad las huellas del cincel de dientes de sierra con que fueron desbastados los bloques de mármol. En la mitad inferior de la fotografía se puede ver la junta entre dos bloques de ésos. En la Edad Media, un violento terremoto, que demolió la mayor parte de lo que permanecía de pie en el templo, desplazó algo los bloques, como se aprecia en el ligero desvío en la alineación de las líneas verticales inscritas.



8. TRES DIBUJOS SUPERPUESTOS distínguense en las líneas grabadas en la pared del templo (*líneas en rojo*). Estos dibujos definen la forma de la basa y del fuste de una columna: dos secciones longitudinales de la columna (una de ellas dibujada acostada) y otra horizontal del fuste de la misma. La sección longitudinal representada en posición vertical está reducida a una escala de un dieciseisavo de su tamaño real; los otros dibujos aparecen a escala natural. También mostramos aquí las juntas de los bloques de mármol de la pared (*líneas en negro*). Cada gráfico sólo pretendía dibujar la mitad del elemento arquitectónico que representaba, pues la otra mitad era simplemente su imagen especular. De la columna acostada sólo vemos el cuarto superior de su sección longitudinal. El arco trazado debajo de su cuerda pone de manifiesto la curvatura casi imperceptible del contorno del fuste. Por otra parte, la sección horizontal y diagramática del fuste ofrece en su interior tres sectores que corresponden al espaciado de las 24 estrías que rodean la circunferencia de la columna. Por último, en el tercero de los dibujos, cada uno de los apretados renglones paralelos representa un pie griego de longitud a lo largo del fuste de la columna. En esta versión abreviada se puede ver con perfecta claridad el contorno convexo de la columna. Sobre la base de este dibujo se podían obtener directamente las dimensiones variables del radio del fuste. Los dos cuadraditos en rojo señalan las áreas que abarcan las fotografías de la figura 7. (Los dibujos son de Alan D. Iselin.)

El más revelador de los planos de la columna es el que hemos considerado en primer lugar: el perfil de la basa de la columna. Sobre el perfil de la basa está el rayado que primero me llamó la atención: las líneas horizontales paralelas separadas por cortos espacios. Estas líneas están trazadas entre dos líneas verticales que representan, respectivamente, la superficie exterior del fuste de la columna y su eje interior.

La curvatura del fuste está representada en este dibujo mediante un arco convexo de poca profundidad, cuyos extremos están enlazados por una cuerda diagonal (en parte borrada por la erosión). El hueco entre el arco y la cuerda correspondiente mide 4,65 centímetros en su punto más ancho. La distancia coincide exactamente con la mayor separación, en el centro del fuste, entre las dos líneas inferiores de la sección grande transversal y vertical. Los radios de los extremos superior e inferior del fuste (que observan una relación de 5 a 6 entre uno y otro, exactamente) concuerdan también en ambos dibujos. Sólo la altura del fuste varía grandemente entre estas dos representaciones de la columna: en la última se ha encogido de 18 metros a poco más de un metro, o sea, un dieciseisavo de su tamaño real. El dibujo, en contraste con todos los demás, experimentó una reducción de escala en su tamaño, pero sólo en su longitud, no en su anchura. Había una razón para ello.

Uno de los refinamientos estructurales de los templos griegos es el suave encorvamiento que se observa no sólo en la imperceptible curvatura del fuste de la columna, como más arriba se ha dicho, sino también en el ligero arco a que se amoldan todas las capas horizontales del templo. Sobre una distancia de 90 metros, el edificio entero —de los cimientos al tejado— "se hincha" hacia arriba. En el punto medio de su eje longitudinal el basamento del templo se encuentra a 11 centímetros por encima del nivel de cada extremo.

A este toque sutil los griegos lo llamaban, con bastante propiedad, éntasis, o tensión. Constituye la expresión más vigorosa de la "actividad" producida por las partes individuales de la estructura, tanto las partes sustentantes de la carga como las que son, de suyo, sustentadas. Era merced a este refinamiento como las partes podían fusionarse en un todo orgánico en vez de estar aisladas unas de otras. Hasta ahora había sido un misterio saber cómo se lograba ese contorno delicadamente encorvado de los fustes.

Tal como se ve en los dibujos del templo de Apolo, el arco seguido por el contorno de la columna debía hincharse hacia afuera, hasta una distancia máxima de 4,65 centímetros de la cuerda recta que conectaba los extremos superior e inferior del fuste de 18 metros. Un arco tan plano, en dimensiones menores, podía dibujarse con un trazo de compás. Pero haciéndolo a su verdadero tamaño hubiera sido imposible trazar la curvatura de la columna, pues el compás tendría que montarse para un radio de casi nueve décimas de kilómetro.

Por consiguiente, lo que se empleó fue una construcción geométrica tan simple como ingeniosa: la altura del fuste se redujo. Para hacerlo así, se equiparó el pie griego, en sus dimensiones reales (29,6 centímetros), con un dáctilo, un dieciseisavo del pie (1,85 centímetros). La curva del éntasis de la columna podía ahora trazarse con un compás que tuviera un radio de aproximadamente 3,2 metros. Dejando las medidas de anchura sin cambiar, era posible obtener las medidas a escala natural de los radios de la sección transversal, sujetos a continuos cambios, en cualquier punto de la columna. La altura correspondiente, a lo largo de la columna, para cada medida de un ra-

dio, podía calcularse, a partir del dibujo a escala, contando las líneas del rayado; la separación entre cada dos de ellas es de un dáctilo y, por tanto, el ancho de cada renglón representaba un pie, siguiendo la longitud de la columna real.

Estos someros bosquejos habrían bastado para guiar a los canteros cuando empezaban a labrar las columnas. Sin embargo, los planos están completados por un segundo juego de otros más detallados al otro lado del *ádyton* y frente a los bosquejos preliminares: una sección de 18 metros de longitud, a lo largo del fuste de la columna, y dos secciones transversales, una de la parte alta y otra de la baja del fuste. En estos dibujos, todo lo que habíase someramente bosquejado en los planos anteriores está dibujado con sumo detalle. Por ejemplo, las secciones horizontales están terminadas hasta los pormenores de todas las estrías. Sólo entonces se sintió el arquitecto satisfecho: los planos de las columnas eran tan perfectos cuanto las columnas estaban destinadas a serlo.

También se han encontrado los planos de construcción de otras partes del templo. Una extensa colección de dibujos parece referirse a la construcción de las paredes del templo, cuyas superficies externas no se alzan estrictamente verticales, sino que se inclinan ligeramente hacia dentro. Otro dibujo ha manifestado representar parte del entablamento sostenido por las columnas.

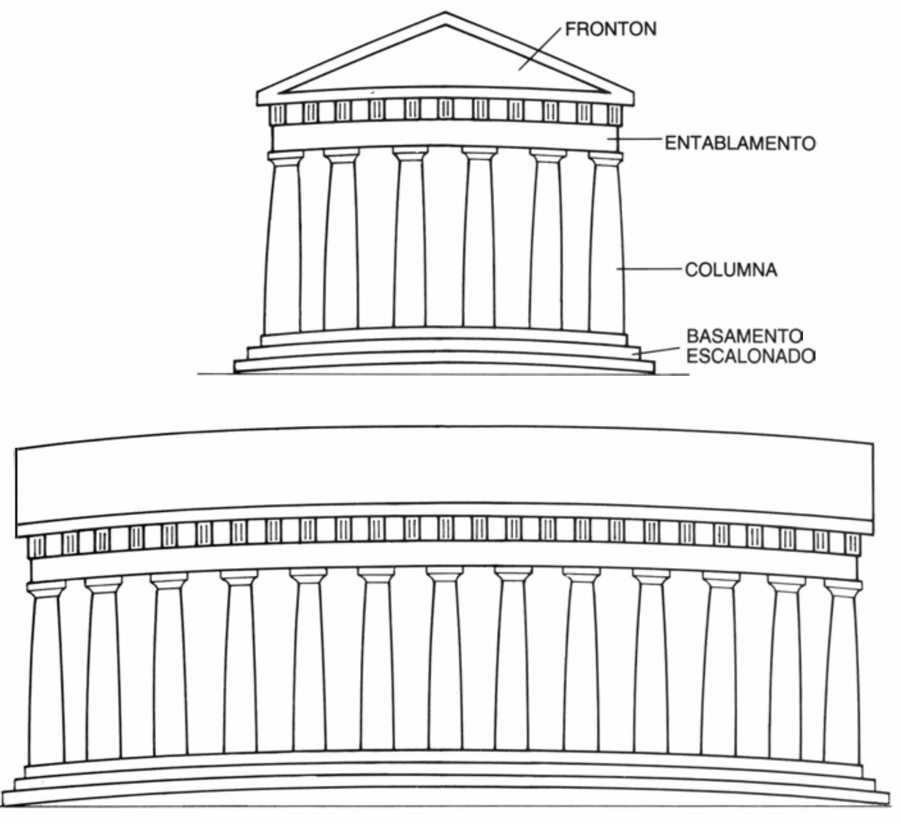
Un dibujo de 12 metros de ancho, que muestra el entablamento y el frontón (el bajo gablete de la fachada) del *naískos*, se descubrió en el muro transversal del fondo del *ádyton*. No constituyó, en cambio, una sorpresa comprobar que el dibujo de la construcción del entablamento correspondía exactamente a lo que en la práctica se había construido. Lo sorprendente fue, sin embargo, la profunda discrepancia entre las dimensiones del frontón dibujado y lo que en realidad se construyó, diferencia que afectaba a toda la edificación. El *naískos* fue edificado unos 2,8 metros más estrecho que lo indicado en los planos, y sus proporciones relativas se vieron así profundamente alteradas. Una libertad de esta magnitud resulta difícil de explicar, incluso a la luz de cuanto sabemos acerca de la arquitectura griega.

Los 200 metros cuadrados de dibujos descubiertos representan una parte grande, pero de ningún modo la totalidad de los planos que deben haberse realizado. Por ejemplo, no se ha encontrado ningún dibujo que muestre el frontón triangular del templo mayor. Esto no es inesperado del todo, puesto que la construcción del frontón principal no llegó a iniciarse nunca; es muy probable que ni siquiera fuese planificada y dibujada en todos sus pormenores. Por otra parte, los planos de las columnas del *naískos*, que habían sido erigidas, se echan también en falta, o por lo menos no se encuentran en los muros del templo.

El juego de planos más crucial que no aparece por ningún lado en los muros es el de los cimientos del edificio entero. Es evidente que debieron ser anteriores a los planos del alzado del templo. Esta omisión no se puede echar en saco roto. ¿Dónde estaban las superficies en que deberían estar dibujados?

Si los griegos dibujaban los planos de las paredes de un templo en las propias paredes, ¿pudieron dibujar los planos del piso en el piso? Con esta pregunta en la cabeza examiné detenidamente las capas de piedra que componen la cimentación del templo. La búsqueda reveló todo un despliegue de líneas finas, grabadas en la superficie de las capas individuales: alineamientos, ejes y otras líneas de construcción. Los detalles del plano del piso habían sido proyectados en las capas mismas que formaban la plataforma escalonada. Cada capa se había echado de acuerdo con los planos, inscritos a su tamaño en la capa precedente. Además, ideas de orden más general para la construcción del templo eran consignadas como sugerencias en una capa, y en caso de ser aceptadas, sus trazos se copiaban en las capas sucesivas. Una vez llegados al estrato superior del basamento escalonado, el arquitecto se encontraba ante el plano definitivo en que las partes más importantes del alzado (los muros y columnas) habían de asentarse cuidadosamente en su forma final y coherente.

La colocación de estas partes estaba reglamentada por una precisa cuadrícula rectangular. Las guías de la cuadrícula se encuentran en el único sitio apropiado: siguiendo la base de la pared que rodea el núcleo del templo. Los indicadores consisten en una serie



9. LIGERAS CURVATURAS se encuentran a lo largo de todos los miembros verticales y horizontales de los templos griegos. En el boceto esquemático que presentamos, las curvaturas se han acentuado en la vista frontal (*arriba*) y en la vista lateral (*abajo*) de uno de tales templos. Las columnas, con disminución ascendente, se hallan también algo "hinchadas" y el templo entero se arquea. Las desviaciones reales de la vertical o de la horizontal estrictas son minúsculas; las columnas del templo de Apolo "se hinchan" hacia afuera un máximo de 4,65 centímetros a lo largo de casi 20 metros; la cimentación "se inflama" un máximo de 11 centímetros sobre un trecho de 90 metros. Y sin embargo, estos refinamientos dan vigor claramente al diseño del templo.

de líneas verticales cortas que, aunque no llamativas, han sido grabadas con una precisión y una exactitud milimétricas.

Con el descubrimiento de estos planos quedaba detectado el eslabón existente entre el plano horizontal y las construcciones verticales dibujadas en las paredes del templo. Ahora tenemos ante nosotros una secuencia ininterrumpida de planos.

No hace falta decir que los arquitectos del templo no esperaron a que estuviesen asentadas las primeras capas de piedra de la cimentación para empezar a contemplar la planta definitiva, ni tardaron en fijar las dimensiones de conjunto que el edificio había de tener antes de levantar los muros del mismo. Los dibujos de las paredes y de otras superficies representan sólo la parte del proceso de planificación que precede inmediatamente a su ejecución: la planificación del solar.

Antes de llegar a esta fase final los planos del edificio debieron haber pasado por varios trazados. Parece probable que el papiro, el pergamino, tablillas blanqueadas, o incluso piedras lisas, se hayan utilizado como materiales de dibujo para estos diseños preliminares. La última de estas posibilidades ha sido confirmada hace poco por mi colega Wolf Koenigs. Mientras éste estudiaba el templo de Atenea en la ciudad de Priene, descubrió que había grabado un apunte a escala del frontón del edificio en la parte inferior de un bloque, que más tarde quedó incorporado a la construcción del mismo.

Koenigs observó también vestigios de dibujos a gran escala en uno de los lienzos murales del templo. Yo he encontrado vestigios de dibujos, grabados de modo similar, en el templo de Artemis en Sardes; y Wolfram Hoepner ha descubierto testimonios de planos de una cámara funeraria que fueron dibujados con tiza roja.

Los dibujos antiguos de Dídyma pueden no ser ya los únicos de que haya constancia en la arquitectura griega, pero son desde luego los más completos. Es una ironía que esta fortuna arquitectónica se haya conservado gracias a que el templo nunca llegó a estar próximo a su terminación.

Mientras que las señales hechas en las superficies horizontales de las piedras desaparecían en su mayor parte al construir encima, los dibujos de las paredes continuaban visibles. Y sin embargo, hacia el final de la construcción, también ellos estaban llamados a ser borrados. Las superficies de las paredes del templo de Apolo no tienen el acabado que se les reservaba, y aún ostentan una delgada capa protectora que no estaba llamada a ser raída hasta la última fase de la construcción. El pulimentado final sólo se puede ver en las esquinas interiores de la pared (donde después no se habría podido conseguir con la perfección deseada). El espesor de la capa protectora mide aproximadamente medio milímetro, lo que viene a ser la profundidad media de las líneas incisas. Si se hubiesen sometido las paredes al pulimento final, los intrincados planos de construcción que habían guiado la edificación del templo hubieran quedado destruidos.

La razón de la aparente falta total de planos de construcción en la arquitectura griega radica sobre todo en una característica que puede calificarse de única. En cada estadio de la construcción, los planos de un edificio se elaboraban y refinaban de continuo. Una vez terminado el edificio, sin embargo, ya no había necesidad de los planos. Este proceso refinador desembocó en creaciones singulares de una belleza intemporal, insuperables en su estilo autodefinidor, clásico.

Lo más gracioso es que, para alcanzar tal perfección, los griegos usaron procedimientos que no diferían esencialmente de los empleados antes por los arquitectos egipcios, utilizados luego por los romanos y más tarde aún por los constructores de la Edad Media.

El lado técnico y práctico del diseño y del dibujo a gran escala en la arquitectura griega difería poco de lo que era evidentemente una práctica común 2000 años antes en Egipto y 1500 años más tarde en Europa central. Se han encontrado planos de construcción grabados en las superficies de las paredes y de los suelos de muchas iglesias góticas (por ejemplo, las catedrales de Chartres, Reims, Brujas, York y Orvieto). Y "líneas de medida y de guía", pintadas o grabadas, se han observado en monumentos primitivos egipcios que pueden considerarse ahora, a la luz de los dibujos de Dídyma, como planos de construcción.

Mi descubrimiento en el templo de Apolo no desvela ningún conocimiento "perdido". Sólo demuestra que, aunque los griegos emplearon métodos corrientes para traducir sus ideas arquitectónicas a monumentos tangibles, alcanzaron extraordinarios y duraderos resultados.

Juegos de ordenador

Búsqueda de una regla invisible que ayude a los radioastrónomos en la medición de nuestro planeta

A. K. Dewdney

Una sencilla regla de un pie de longitud, provista de 13 marcas de pulgada, permite medir 12 longitudes distintas. ¿Podrá mejorarse ese conocido instrumento, de modo que permita medir más longitudes que marcas haya en la regla? La respuesta es afirmativa: se pueden eliminar de la regla todas las marcas, menos cinco, y seguirán midiéndose todavía 10 longitudes diferentes. Cada una de esas longitudes corresponderá a la distancia entre dos marcas, igual a la diferencia de los valores enteros con que se hayan rotulado las marcas. Cabe incluso alcanzar ese mismo resultado con reglas de 11 pulgadas. Los lectores que, intrigados por ese ejercicio, acaben resolviéndolo con éxito habrán creado una regla de Golomb.

La búsqueda de tales reglas es una tarea absorbente, en la que los ordenadores pueden prestarnos concurso. Lo que eleva el proyecto de mera curiosidad a rompecabezas de primera es que la necesidad de reglas de Golomb aflora en todo un abanico de disciplinas científicas y técnicas.

La invención de esos artilugios se debe a Solomon W. Golomb, profesor de matemáticas e ingeniería eléctrica de la Universidad de California del Sur. A lo largo de dos decenios, Golomb y un puñado de colegas han buscado las reglas y estudiado sus propiedades. Las reglas "golombinas" pueden tener aplicaciones en teoría de codificación, en el análisis cristalográfico por rayos X, en la distribución de circuitos y en radioastronomía.

Entre los investigadores cuyo trabajo se ensancha y facilita con las reglas de Golomb se cuenta Douglas S. Robertson, geofísico que trabaja para el Servicio Geodésico de la Administración Oceánica y Atmosférica de los Estados Unidos en Rockville, Maryland, quien se vale de un método conocido por interferometría de muy larga base, pero no para cartografiar fuentes astronómicas de radio-emisión, sino con la finalidad de lograr mediciones finamente sintonizadas de nuestra Tierra. Habiendo sido buscador de reglas golombinas durante años, Robertson no duda en suplicar a los lectores que amplíen la búsqueda. El resultado puede contribuir tanto a una determinación más precisa del tamaño, forma y movimiento de nuestro planeta como a hacer más apasionante el tiempo dedicado a ello.

Antes de tratar de responder a la llamada de Robertson, vale la pena dominar los principios subyacentes a las reglas de Golomb. Existen reglas de todos los tamaños, pero sólo se conocen las más pequeñas. Cabe describir las tres primeras, algo abstractamente, mediante tres series de números:

0, 1
0, 1, 3
0, 1, 4, 6

Para construir la regla marcaremos el extremo izquierdo de un listón de n unidades de largo con el menor de los números de la sucesión (o sea, con 0). Grabemos en su extremo derecho el número máximo (n), que puede ser 1, 3 o 6. Los enteros que intervengan deberán ir acompañando a marcas situadas a intervalos adecuados, contados desde el extremo 0 de la regla [*véase la figura 1*].

Una forma sencilla de verificar qué distancias de 0 a n resultan medibles con esta regla consiste en trazar el diagrama de distancias de la regla. En correspondencia con cada uno de los enteros que figuren en la regla, marquemos un punto en una hoja de papel, y rotulémoslo con ese entero. Seguidamente, unamos cada par de puntos por un trazo recto, rotulado con la diferencia entre los enteros que ocupan sus extremos. Se dice que la regla de Golomb es perfecta cuando cada una de las distancias intermedias entre 0 y n figura una vez, y solamente una, en el diagrama de distancias. Las tres reglas de la ilustración son perfectas todas ellas, hecho que puede verificarse sin más que echar una ojeada a sus diagramas de distancias. En ellos, ninguna distancia aparece más de una vez, mientras que se encuentra presente toda posible distancia entre 0 y n.

Entre las reglas de Golomb, la perfección es rara. En realidad, las únicas reglas perfectas que existen son las tres recién descritas. Para valores de n mayores que 6, la imperfección se manifiesta de una de estas dos formas: o bien hay distancias que aparecen más de una vez, o bien las hay que no aparecen.

Es esta una cruel realidad, que es preciso afrontar nada más empezar nuestra búsqueda de reglas mayores (y mejores). ¿Cómo sabemos que no existen reglas perfectas más grandes? Golomb ha proporcionado una demostración tan breve como encantadora.

Golomb sugiere que no nos fijemos en las divisiones de la regla, sino en las distancias que median entre ellas. Si la regla es perfecta, todas las longitudes 1, 2, 3..., hasta $m-1$, habrán de venir dadas (en cierto orden) por distancias entre marcas consecutivas (siendo m el número de marcas). Golomb pregunta: "¿Dónde se encuentra el intervalo de longitud unidad?" Si estuviera adyacente a un intervalo cuya longitud fuera menor que $m-1$, los dos intervalos, conjuntamente, definirían una longitud menor que m. Ahora, esta distancia tiene que estar señalada ya en algún otro lugar, puesto que todas las distancias, desde 1 a $m-1$, corresponden a intervalos entre marcas consecutivas. Tal razonamiento nos obliga a aceptar la conclusión, un tanto sorprendente, de que el intervalo de longitud unidad ha de encontrarse contiguo al de longitud $m-1$. Además, del otro lado del intervalo unitario no puede quedar espacio alguno. Se encuentra, por consiguiente, en un extremo de la regla.

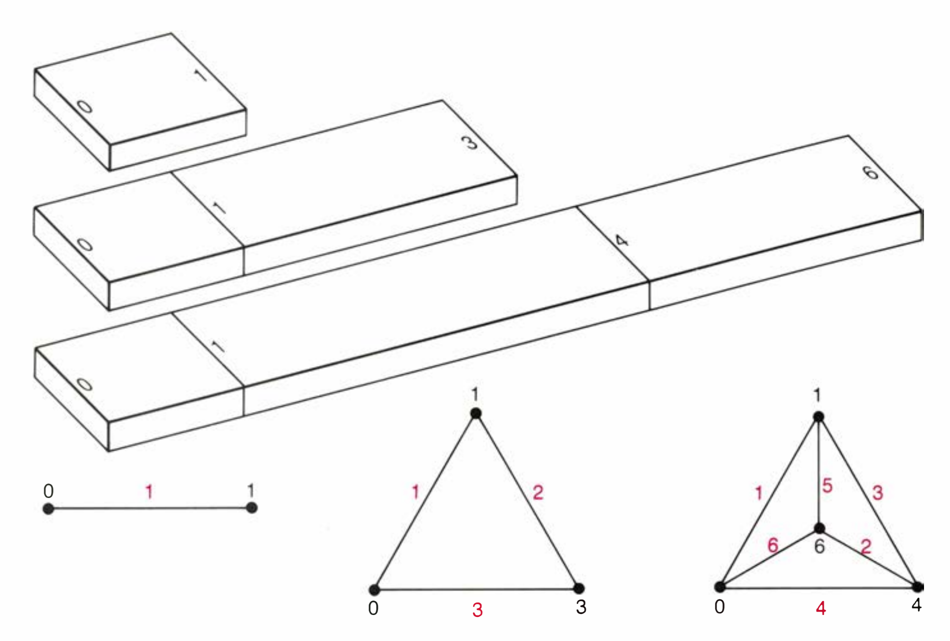
El razonamiento anterior equivale a una vuelta de manivela en lo que ciertos matemáticos denominan "máquina de churros" (y que otros continúan llamando razonamiento por recurrencia).

"¿Dónde se encuentra el intervalo de dos unidades?", pregunta Golomb ahora. Se vuelve a hacer girar la manivela: si el intervalo de longitud dos se encontrase adyacente a un intervalo de longitud menor que $m-2$, entre ambos sumarían una distancia que ya se ha dado en otro lugar de la regla. Esta vez no podemos concluir que el espacio de dos unidades se encuentre adyacente al de $m-2$ unidades. Su longitud conjunta, m, está medida ya por los dos intervalos de longitudes 1 y $(m-1)$, situados en uno de los extremos de la regla.

La máquina de churros se detiene, generando la conclusión de que tan sólo hay un espacio al que pueda estar adyacente el intervalo de dos unidades, a saber, aquel cuya longitud sea de $m-1$ unidades. Dado que ninguna regla tiene más de dos extremos, una regla perfecta tiene a lo sumo tres intervalos, de longitudes 1, $m-1$ y 2. La demostración queda terminada cuando nos fijamos en que tres intervalos requieren cuatro marcas: $m=4$. Así pues, los intervalos miden 1, 3 y 2; las correspondientes marcas están situadas en 0, 1, 4 y 6.

Encarado con una total carencia de reglas perfectas que tengan más de cuatro marcas, el matemático suprime pérdidas construyendo una nueva definición. Queda así de manifiesto lo que se ha dado en llamar "síndrome del casi óptimo": lo más parecido a una regla perfecta de cinco marcas podría ser una regla que contuviera solamente una vez cada distancia, aunque no contuviera todas las distancias que una regla perfecta de esa longitud habría de tener. Dado que tal condición es fácil de satisfacer haciendo la regla suficientemente larga, se añade una cláusula adicional: de entre todas las reglas provistas de cinco marcas que contienen a lo sumo una vez cada distancia, determínese la más corta. Tal regla será llamada regla de Golomb de orden cinco. Las reglas golombinas de órdenes superiores se definen de igual modo. Dado que esta definición incluye como casos particulares a las tres reglas primeras, sirve como puente con que salvar la discontinuidad en la perfección al trascender las cuatro marcas.

Herbert Taylor, colega de Golomb, ha resumido en una tabla el estado actual de la información de que se dispone en lo tocante a reglas golombinas [*véase la figura 2*]. En la zona que va desde las dos a las 24 marcas se tiene tanto conocimiento cierto como conjetural acerca del tamaño de las reglas de Golomb. Lo que yo llamo zona de perfección se extiende desde dos hasta cuatro marcas. A partir de ahí, la zona de conocimiento abarca las reglas golombinas que tienen hasta 13 marcas. Se sabe que todas las reglas de esta zona son mínimas. Es decir, en ninguno de esos casos existe una regla de Golomb más corta que tenga el mismo número de marcas. Una regla de Golomb de cinco marcas tiene longitud 11. Una regla de Golomb de 13 marcas tiene longitud 106.



1. Las tres reglas golombinas perfectas y sus diagramas de distancias

A partir de las 13 marcas comienza la que llamo "zona de penumbra". Bautizada con el honroso nombre de "zona de investigación", contiene sólo reglas de las que todavía no se sabe que sean de Golomb. Para cada uno de los números de marcas existe una regla de longitud dada en la tabla. Pero puede que haya reglas más cortas. Existe, en efecto, una fórmula que da un límite inferior para esas longitudes. El hiato, cada vez mayor, que se abre entre los valores dados por la fórmula y las reglas descubiertas hasta el momento atestigua, bien una debilidad de la fórmula, bien que las reglas obtenidas son cada vez más imperfectas.

Robertson es el responsable de haber ampliado la zona de conocimiento con las reglas de 13 marcas. En una pasada de ordenador, que duró todo un mes, sin interrupción, su programa exploró exhaustivamente el universo de todas las potenciales reglas de Golomb provistas de 13 marcas, y descubrió la más corta de todas. Tal vez sean muy pocos los lectores interesados en hallar reglas golombinas de 14 o más marcas si los tiempos de funcionamiento necesarios son todavía más largos.

En lugar de esas investigaciones exhaustivas, parece razonable emprender ciertas tácticas de sondeo, a modo de incursiones informáticas en la zona de investigación, que prometan algún botín, en forma de reglas mejores que las conocidas. En cualquiera de tales intentos resulta fundamental un programa llamado INSPECTOR, que toma a su cargo la siguiente tarea: dada una tabla de enteros, ¿cuál es el modo más eficaz de determinar si las diferencias entre ellos son todas distintas? Un enfoque simplón consistiría en generar todos los posibles pares de enteros, y recoger todas sus diferencias en otra tabla. Seguidamente, el programa examinaría esta segunda lista, buscando repeticiones, recurriendo a un algoritmo desgarbado y lento.

Raras veces una forma más rápida de ejecutar una tarea se basa en un programa más corto; sin embargo, he aquí un caso. Dado que se supone que las diferencias propiamente dichas son todas distintas, podemos utilizarlas como direcciones de una tabla especial llamada *control*. En principio, en *control* tan sólo se almacenan ceros. Cada vez que se calcula una nueva diferencia, el valor almacenado en la correspondiente dirección se cambia de 0 a 1. De este modo, conforme INSPECTOR prosigue su cómputo, puede encontrarse con un 1 almacenado ya en una dirección determinada, lo que conlleva que la "nueva" diferencia ha aparecido con anterioridad. En tal caso la regla no sería una regla golombina, porque no resiste el criterio fundamental de "golombicidad", a saber, que cada distancia se presente una sola vez.

	NUMERO DE MARCAS	MINIMA REGLA CONOCIDA	COTA INFERIOR
ZONA DE PERFECCION	2	1	
	3	3	
	4	6	
ZONA DE CONOCIMIENTO	5	11	
	6	17	
	7	25	
	8	34	
	9	44	
	10	55	
	11	72	
	12	85	
	13	106	
ZONA DE INVESTIGACION	14	127	114
	15	155	133
	16	179	154
	17	199	177
	18	216	201
	19	246	227
	20	283	254
	21	333	283
	22	358	314
	23	372	346
	24	425	380

2. Longitudes de las reglas golombinas y cuasi-golombinas

La técnica de utilizar las diferencias a modo de direcciones de memoria constituye una forma primitiva de lo que en ciencias de cómputo y en programación se conoce por "condensación" (*hashing*). En muchos problemas de recuperación de información, el método de condensación es el modo más rápido de que dispone el ordenador para recobrar un archivo.

Para quienes tengan necesidad de mayores detalles, he aquí, en esencia, el programa INSPECTOR. Mediante dos bucles encajados se generan todos los posibles pares de enteros pertenecientes a la tabla de partida. Si el primer bucle genera i y el segundo j, el programa calcula el valor absoluto de la diferencia y la almacena en una variable llamada *dif*. En el paso siguiente, INSPECTOR utiliza el valor de *dif* como si fuera una especie de código de condensación. En lenguaje algorítmico se podría escribir lo siguiente:

si *control* (*dif*) = 1
  entonces dar como resultado "Regla no golombina" y concluir
  en otro caso, control(*dif*) ← 1.

Si el programa nunca dice "Regla no golombina", la regla habrá aprobado el primer examen. Ahora bien, ¿cuál es su lontitud? Muchos métodos nos la pueden dar.

Cabe utilizar INSPECTOR en el modo "en solitario". Ya veo al lector pegado al teclado, contemplando las evoluciones del programa. Ha decidido adentrarse en la zona de investigación, en el nivel de 14 marcas, a la búsqueda de una regla de menos de 127 unidades, que es la mejor de las reglas conocidas. El lector, que por el mal tiempo vuela a ciegas, va con el rumbo perdido. Acaba de someter una sucesión de 14 marcas al análisis del ordenador. El máximo de los enteros de la regla es 124; su excitación es casi insoportable cuando la pantalla del monitor cobra vida: "Enhorabuena, ese conjunto de marcas es correcto". Al programar tal mensaje, el lector hizo promesa de someter a INSPECTOR sólo conjuntos potencialmente capaces de batir el récord.

Quizás el lector haya encontrado ese conjunto plusmarquista siguiendo los consejos de Golomb: explorar sólo las reglas en las que el intervalo máximo aparece en el centro. En tales reglas, las separaciones van acortándose hacia los extremos, si bien el valor de la mengua queda a juicio del lector. Golomb nos asegura que muchas reglas buenas, cuando no forzosamente las mejores, obedecen a tal norma.

Puede modificarse INSPECTOR de modo que se ajuste a un estilo de indagación más aproximativo. En INSPECTOR PASO A PASO, los enteros van comunicándose al ordenador de uno en uno. En cuanto se le introduce un dato, el programa genera las diferencias entre el entero recién introducido y los que ya se encontraban almacenados. En realidad, INSPECTOR PASO A PASO no es sino una versión de INSPECTOR en la que el bucle exterior se sustituye por una sentencia de entrada de datos. La secuencia tiene éxito si llega a digerirse el último entero sin que el programa suelte "no golombina".

Finalmente, puede incorporarse el programa INSPECTOR PASO A PASO en una búsqueda automatizada, como la emprendida por Robertson. Su programa (al que bien podríamos llamar EXHAUSTACION, dado su carácter) genera nuevas reglas a base de ir añadiendo sistemáticamente un espacio cada vez. Después de cada adición, INSPECTOR PASO A PASO decide si la regla que en ese momento se encuentra en construcción es válida.

Robertson construyó su programa visualizando una regla a la que se añaden nuevos espacios (y nuevas marcas, por tanto) de izquierda a derecha. Los lectores que hayan seguido el sendero de nuestra prosa, y atravesado el puerto que supone el razonamiento de Golomb (el cual demuestra la inexistencia de reglas perfectas) recordará que los intervalos que tienen que aparecer (en un cierto orden) son de longitudes 1, 2,..., $m-1$. Si bien ello solamente es cierto para las reglas perfectas; para las demás reglas de Golomb vale algo similar. En las que no llegan a perfectas se encuentran, sumando intervalos consecutivos en un cierto orden, casi todas las longitudes de 1 a $m-1$. No obstante, en tales reglas pueden existir intervalos de longitudes mayores aún que $m-1$.

Robertson engendra nuevos intervalos procediendo paso a paso. Los conserva en una tabla de valores que llamaremos (adecuadamente, me parece), *intervalos*. El programa EXHAUSTACION va recorriendo la tabla, añadiendo un intervalo tras otro. Como es natural, existen ciertos criterios sencillos que facilitan la labor de EXHAUSTACION. Uno de ellos consiste en asegurarse de que cuando se genere un nuevo intervalo, éste no figure ya en *intervalos*. Un segundo criterio consiste en sumar las longitudes de todos los intervalos que componen la regla en construcción, para asegurarse de que la longitud total no excede la actualmente tenida por mínima.

En su funcionamiento, el programa EXHAUSTACION parece ansioso por hallar reglas. Comienza por hacer igual a uno el primer elemento de *intervalos* y añade unidades al segundo intervalo, a fin de que sea distinto del primero. A continuación añade unidades al tercer intervalo, a fin de que su diferencia no sólo sea distinta de las dos primeras distancias, sino que satisfaga también la condición impuesta por INSPECTOR PASO A PASO, a saber, que todas las distancias contenidas en la regla sean diferentes. Cada vez que de este modo queda admitido un nuevo elemento en la tabla de intervalos, EXHAUSTACION suma la totalidad de las longitudes de la tabla, y compara la suma con la menor de las reglas hasta ahora conocidas. Si la suma es menor, el programa continúa con la siguiente entrada. Si no es menor, EXHAUSTACION regresa al elemento anterior, y continúa añadiéndole intervalos.

El programa de Robertson funcionará algo más rápido si como primer elemento de *intervalos* se toma 2 en lugar de 1. En efecto, los períodos de funcionamiento de un día quedan abreviados en algunas horas. Quizá los lectores encuentren de su agrado razonar por qué la búsqueda continúa siendo exhaustiva.

No cabe duda de que la eficacia de un programa de exploración exhaustiva depende de la inclusión de nuevos criterios y de métodos heurísticos. La imposición de nuevas limitaciones a los valores que pueden tomar los diversos elementos de la tabla *intervalos* permitiría mejorar mucho el rendimiento. Tal vez exista un procedimiento incremental que utilice una gama de valores mucho más estrecha. El procesamiento de la tabla tiene afinidad con el recuento, que se alcanza mucho antes si se logra reducir el número de posibilidades correspondientes a cada elemento de intervalos. En cualquier caso, los lectores que descubrieron castorcitos afanosos, nuevos cañones lanzapatines y demás frutos de la investigación serán sin duda capaces de abrirse paso en el campo de investigación golombina.

Las nuevas reglas que descubran los lectores se publicarán en un artículo futuro.

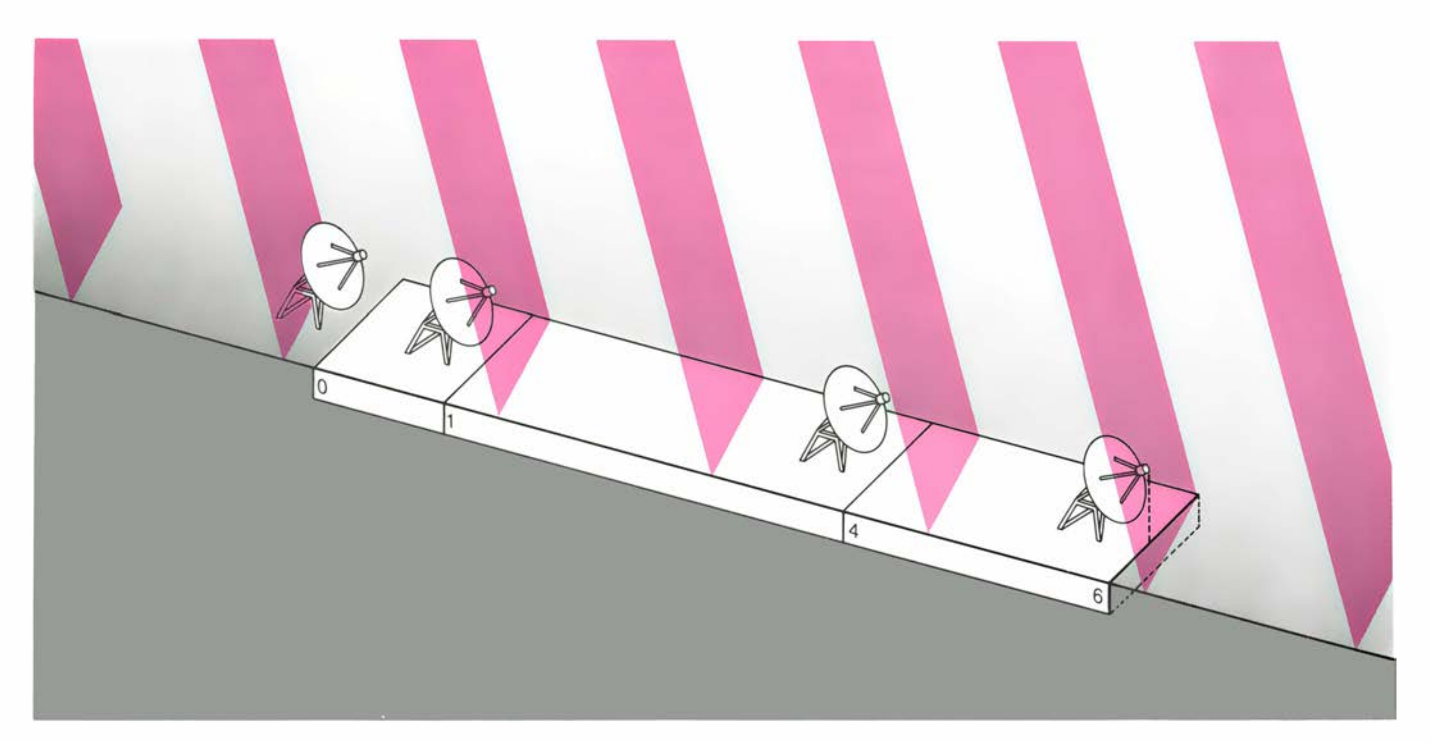
Golomb ha ofrecido un premio de 100 dólares al primero que encuentre dos reglas distintas, de más de seis marcas, que midan, sin embargo, el mismo conjunto de distancias diferentes. Las reglas imagen simétrica una de otra no se consideran distintas.

De obtenerse un resultado positivo, las campanas habrán de doblar a muerto por un "teorema" propuesto en 1939 por Sophie Piccard, matemática suiza. El teorema de Piccard enuncia que dos reglas capaces de medir iguales conjuntos de distancias tienen que ser reglas iguales. El teorema fue abrazado por los especialistas en análisis cristalográfico por rayos X, porque les ayudaba a resolver las ambigüedades de los patrones de difracción. Desafortunadamente, el teorema falla para numerosas parejas de reglas de seis marcas. Quizá sea verdadero para todas las reglas de orden superior. Los lectores pueden aspirar al premio de Golomb sin necesidad de aventurarse en la zona de investigación. Es cuestión que puede abordarse para reglas marcadas con tan sólo siete divisiones.

¿Qué relación guarda todo ello con la posibilidad de ayudar a Robertson? En ocasiones se usan en radioastronomía reglas de Golomb para la resolución de fuentes de radioemisión muy alejadas, y para la medición de las dimensiones de nuestro propio planeta. En el primer caso se sitúa cierto número de antenas a lo largo de una línea recta de varios kilómetros de longitud. Las posiciones de las antenas corresponden a las marcas de una regla de Golomb [*véase la figura 3*]. Para localizar una fuente radioemisora distante es esencial determinar el ángulo que tiende la línea base de las antenas con la dirección de los frentes de onda procedentes de la radiofuente. Las antenas se sintonizan todas a una determinada longitud de onda. El momento exacto en que llega cada onda de la señal incidente a cada una de las antenas se determina analizando la cinta donde se registra la señal. El número total de longitudes de onda que hay entre un par de antenas dado se llama diferencia total de fase. De ordinario, la diferencia total de fase se compone de una parte entera y de una parte decimal, llamada diferencia de fase. Si pudiera reconstruirse la diferencia total de fase, el ángulo buscado, que es el ángulo que forman la fuente y la línea de base, se derivaría fácilmente a partir de la longitud de onda en que se efectúa la observación y de la velocidad de la luz, c. Sin embargo, cada par de antenas tan sólo permite saber la diferencia de fase, pero no la diferencia total de fase.

A decir verdad, es por análisis armónico como se reconstruye la diferencia total de fase a partir de los registros de muchos pares de antenas. Pero si la distancia entre un determinado par de antenas es la misma, o casi, que la distancia que media entre otro par, uno y otro suministrarán la misma información sobre diferencias de fase. En este caso, la redundancia de información produce pérdidas en la misma. La precisión del cálculo del ángulo de la fuente es máxima si cada uno de los pares de antenas registra una diferencia de fase diferente; y ese requisito se logra, en efecto, situando las antenas en las marcas que definen una regla de Golomb.

Otro método de localización de radiofuentes distantes consiste en utilizar solamente dos receptores, que van simultáneamente revisando cada uno un



3. Con pares de antenas radiotelescópicas situadas sobre una regla de Golomb se determinan las diferencias de fase de las señales incidentes

conjunto completo de longitudes de onda. Al observar con dos antenas una fuente distante, a varias longitudes de onda diferentes, se genera la misma información acerca de la diferencia total de fase entre las antenas que con varias antenas sintonizadas a una misma longitud de onda. Vuelve aquí a amenazarnos el problema de la redundancia. No debe haber, por así decir, dos pares de longitudes de onda exploratorias separadas por una misma distancia. La regla de Golomb resulta ahora invisible, pero sigue presente.

Robertson se vale de esta segunda técnica, pero no para cartografiar las radiofuentes, sino para localizar las antenas propiamente dichas. Para sus fines no le basta saber que la segunda antena se encuentra en Westford, Massachusetts, sino que necesita conocer su posición con un error de pocos centímetros. Puede localizarla con esa precisión valiéndose de una radiofuente puntual muy alejada, como un quásar. Localizar un único punto de la superficie terrestre con respecto a una radiofuente muy distante equivale a la determinación precisa de variables terrestres de tanta importancia como el diámetro, la orientación del eje de rotación y la duración del día. Cuando las precisiones son del orden de los centímetros y los microsegundos, dichas variables lo son verdaderamente, y señalan variaciones anuales, estacionales, e incluso meteorológicas, repletas de significado en ocasiones, y misteriosas en muchas otras.

En el artículo de "Juegos de ordenador" del mes de noviembre se describía una enfermedad benigna de las terminales, llamada CANGREJOS. Ocasional e inesperadamente, hordas de crustáceos marinos descienden hambrientos por las terminales Blit de los científicos de los Laboratorios AT&T Bell de Murray Hill, Nueva Jersey.

Prácticamente todas las características del ambiente de multiprogramación de las terminales Blit se fundan en una importante instrucción llamada *bitblt* (transferencia de bloques periféricos). La instrucción *bitblt* transfiere un conjunto rectangular de posiciones de memoria de una a otra zona de la memoria de la terminal. Dos de tales conjuntos rectangulares pueden combinarse mediante diversos operadores lógicos antes de efectuar la transferencia.

Se retó a los lectores a resolver dos intrigantes rompecabezas: uno, borrar de la pantalla una cierta figura; y otro, hacerla girar 90 grados. Resulta fácil resolver el primero. Para borrar una figura que ocupa un conjunto rectangular se combina el conjunto consigo mismo mediante el operador OEX. El resultado consta exclusivamente de ceros. Basta un único mandato de *bitblt* para remplazar la figura primitiva por una pantalla en blanco.

Hacer girar una figura es mucho más difícil que borrarla. Hasta la fecha, tan sólo un lector ha remitido una solución viable y práctica. Thomas Witelski, estudiante de preuniversitario, de Nueva York, ha dado con un procedimiento para girar una figura de $n \times n$ en una pantalla de $2n \times 2n$. Valiéndose de sólo $3n-1$ operaciones *bitblt* recorre $n-1$ hileras, copia una subimagen de $(n-1)\times(n-1)$, se mueve $n-1$ columnas, copia otra subimagen y avanza $n-1$ filas más. El método de Witelski es más rápido que el algoritmo de rotación convencional, que requiere $4n-2$ operaciones.

He recibido lastimeros sollozos de lectores que, deseando ver el programa MANDELZOOM en su ordenador personal, no se sienten capaces de escribir su propia versión del programa. Siempre que los lectores me escriben solicitándome listados, o discos, con los programas que describo, siento la tentación de complacerles. Sin embargo, tardaría demasiado en atender todos los encargos. Además, el tiempo que pueda invertir el lector en aprender los pocos elementos de programación necesarios para la mayoría de los proyectos que aquí expongo es tiempo bien gastado. Al profesor que hay en mí le complace el argumento.

Sin embargo, igual que algunos lectores tienen dificultades para iniciarse en las bellas artes de la programación, otros, más expertos, sienten vivos deseos de compartir sus logros. Se me ha pasado por las mientes que con un mínimo de burocracia podría ponerse en práctica una especie de sistema de camaradería informática. Serviría para establecer contacto entre dos clases de personas: lectores que suponen que van a necesitar ayuda para realizar los programas de "Juegos de ordenador" y lectores dispuestos a prestarla. ¿Me equivoco al suponer que habrá suficiente abundancia de estos últimos? Envíenme una tarjeta postal con su nombre, dirección postal completa, número de teléfono y categoría (novato o consejero). Los lectores dispuestos a dar consejos deberían especificar el número máximo de noveles a quienes se sienten capaces de auxiliar. Por mi parte, procuraré establecer enlaces entre personas situadas a distancias geográficas razonables.

Taller y laboratorio

Calidoscopios equipados con diodos de destello y condensadores ópticos

Jearl Walker

De un calidoscopio diríase que es un instrumento mágico que transforma pequeños fragmentos de vidrio y plástico de color en conjuntos, geométricamente precisos, de formas y colores. Considerado hasta hace poco mero juguete, nuevos diseños y originales e ingeniosas aplicaciones de los principios de la óptica lo han llevado al terreno del arte. Hay calidoscopios modernos valorados en millares de dólares.

Este instrumento lo inventó en 1816 David Brewster, físico británico recordado por sus estudios acerca de la polarización de la luz. El calidoscopio clásico consiste en un tubo que encierra dos o tres superficies reflectoras, tales como espejos o tiras metálicas brillantes, que se extienden a lo largo de todo su interior. Los reflectores pueden formar tanto una V como un triángulo. Al mirar por un extremo del tubo, habitualmente a través de una pequeña abertura, la visual se apunta para que pase por entre los reflectores hacia el extremo opuesto.

En ese otro extremo del tubo se colocan diversos materiales. Entre éstos son corrientes trocitos de plástico o vidrio brillantemente coloreados, así como clips sujetapapeles, alfileres y otras piezas metálicas. En el calidoscopio, parte de lo que se observa es una visión directa de esos materiales. Esta visión directa tiene, en los calidoscopios de dos espejos, la forma de una porción de tarta. En los de tres espejos esa forma es triangular. Pero en ambos casos se ven, asimismo, otras imágenes que son el reflejo de la visión directa. Mucho he cavilado en torno a tales imágenes. ¿Por qué algunos calidoscopios producen sólo algunas, mientras que otros dan centenares? Supuse que la diferencia debía estar en la calidad de los espejos, pero me equivocaba.

En numerosos calidoscopios actuales se incorporan variantes de los diseños tradicionales. En algunos, los objetos del extremo opuesto del tubo se sumergen en un aceite viscoso y transparente para que se muevan despacio al rotar el tubo. Está también muy extendido el uso de ruedas. Estas se montan en un eje que sobresale del tubo. A veces se ponen dos ruedas, que pueden construirse con algún plástico transparente que contenga alas de mariposa o trozos de minerales de colores finamente rebanados. Otras veces la rueda puede hacerse de vidrio o plástico coloreado y traslúcido, combinado de manera que se asemeje a una vidriera de colores. La idea es hacer girar la rueda a la vez que se mira por el calidoscopio. En otros modelos se agregan cajas de música, y cuando se da cuerda al instrumento, el motivo coloreado del otro extremo rota durante varios minutos acompañado de una melodía.

Otra variante se basa en la polarización de la luz. Aquí el extremo opuesto está cubierto de un filtro polarizador y junto a él, por dentro, se colocan tiras de envoltura plástica tensadas. Encima del orificio de observación se monta otro filtro polarizador. De este modo se consigue que el plástico incoloro aparezca brillantemente coloreado merced a la disposición de los filtros. Los colores dependen del espesor de las tiras y de la dirección en que éstas quedaron tensadas.

Hay una versión de este calidoscopio en que las reflexiones se generan mediante una chapa metálica brillante previamente arrollada en forma de cilindro, uno de cuyos extremos se conforma en espiral, que luego se deja suelto dentro del tubo. Las imágenes reflejadas no son reproducciones nítidas de la visión directa, sino franjas de color difuminadas y alargadas. La superficie reflectora de uno de mis calidoscopios la constituye el interior metálico del tubo, fuertemente pulimentado, lo cual origina anillos concéntricos en torno a la visión directa.

En otro de mis calidoscopios los reflectores son de plástico verde claro. Estos producen imágenes especulares, pero su visión se perturba con la luz que se cuela por los costados transparentes del instrumento. Para mejorar la visibilidad, lo cubro con una tela.

Hoy en día un gran número de calidoscopios tienen el extremo opuesto abierto, por lo cual lo que ofrecen es una imagen transformada y cambiante de la zona a la que se apunta el instrumento. De éstos, algunos tienen una lente convergente encolada en el extremo opuesto, con lo que se consigue abarcar un mayor campo de visión. Entre mis calidoscopios hay uno que tiene una lente semiesférica en el extremo con la cara plana hacia adentro. Entre esta lente y los espejos tiene una pequeña bola esférica de plástico. La mayor parte de la visión directa se compone de la escena presente ante el instrumento, concentrada por la lente semiesférica, pero también se ve la misma escena concentrada aún más por la bola esférica.

Mis dos calidoscopios favoritos los he dejado para el final. Uno procede de Kaleidoscopes by Peach, de Austin, Texas. Peach Reynolds, accionista de la compañía, ha sido uno de los iniciadores del desarrollo de los modernos calidoscopios. Una versión de su instrumento constituye el primer calidoscopio "de alta tecnología", en el que los tradicionales trocitos de vidrio y plástico de colores se han reemplazado por un conjunto de diodos emisores de luz que gobierna un transductor accionado por sonido. Cuando este transductor intercepta ondas sonoras, enciende brevemente los diodos coloreados. Para observar las distintas partes del conjunto hay que girar el extremo del tubo. Encuentro muy entretenido sentarme ante mi equipo estéreo mientras los diodos destellan al compás de la música.

Mi otro calidoscopio preferido lo fabrican Timothy Grannis y Lazarowski, de Prism Design, en Essex Junction, Vermont. En éste, una lente montada en el extremo comprime la escena exterior. En ambos extremos los espejos forman triángulos equiláteros. La originalidad estriba en que los espejos se ensanchan hacia el observador, de modo que el triángulo del extremo opuesto es menor que el triángulo del extremo próximo.

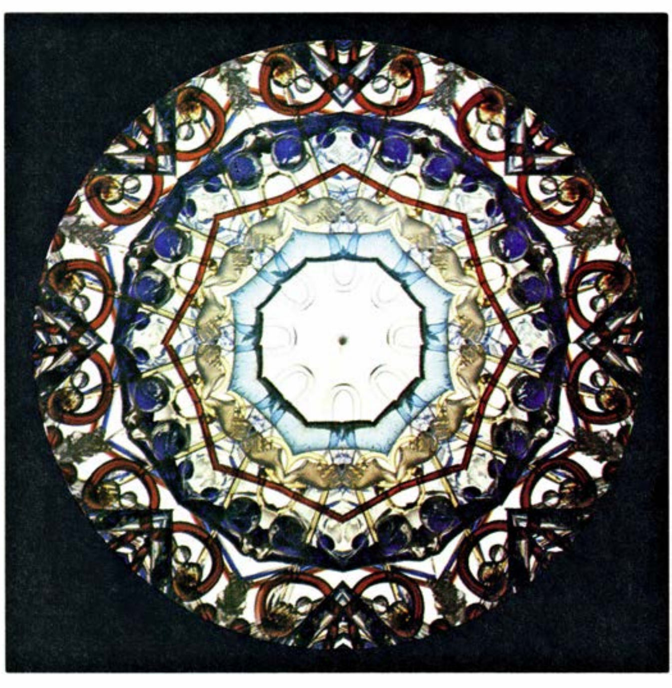
Asombra el conjunto resultante de imágenes reflejadas. Estas no se distribuyen, como en otros calidoscopios, en un plano que atraviesa la visión directa. Lo que hacen es formar una esfera geodésica compuesta de porciones triangulares. Es desconcertante la ilusión de que la figura se curva tridimensionalmente. Esta ilusión la realzan unos hi-

los delgados y brillantes que parecen proyectarse radialmente desde la esfera hacia un fondo luminoso e indistinto. Al apuntar este calidoscopio a una escena callejera, o a una pantalla de televisión en funcionamiento, la esfera estalla en un movimiento frenético, con cada una de sus porciones mostrando algo diferente.

¿Cómo se crean tan bellas figuras en los calidoscopios clásicos y modernos? Yo ya sabía que intervenían reflexiones, pero no comprendía de qué modo se ajustaban tan limpiamente a un orden geométrico. Tras examinar varios de mis calidoscopios seguí igual de intrigado, puesto que resultaban demasiado pequeños para aclarar nada. Dispuesto a construirme un calidoscopio mucho mayor, compré varios espejos planos de 1,3 metros de largo por 0,3 metros de alto.

Por lo corriente estos espejos baratos proporcionan unas imágenes pobres. En ellos, la mayor parte de la reflexión proviene de un revestimiento metálico brillante que cubre el dorso del vidrio. La superficie delantera refleja menos. Ambas reflexiones producen unas imágenes con los bordes borrosos. No obstante, mis espejos me dieron unas imágenes notablemente claras.

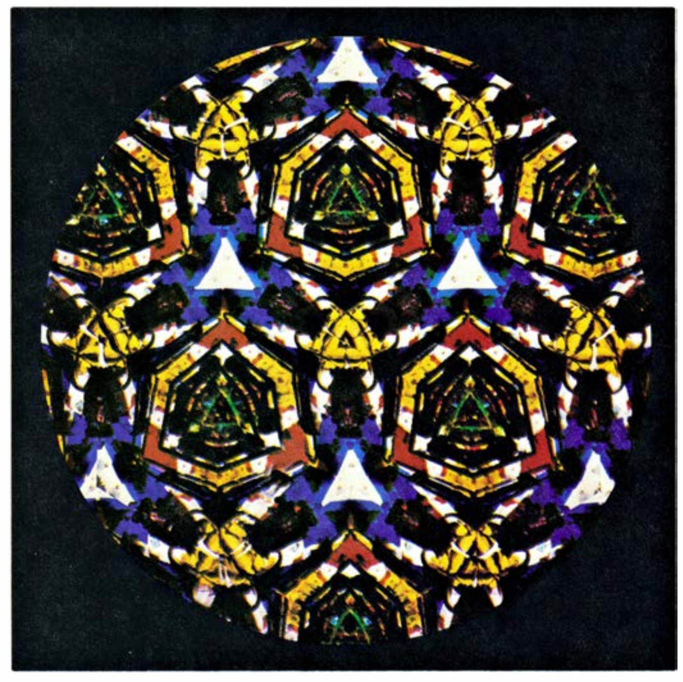
Coloqué en posición vertical dos de estos espejos y los uní con cinta adhesiva de modo que formaran una V, con las superficies reflectoras en el interior. Así, aquella especie de bisagra de cinta adhesiva me permitiría variar fácilmente el ángulo entre los espejos.



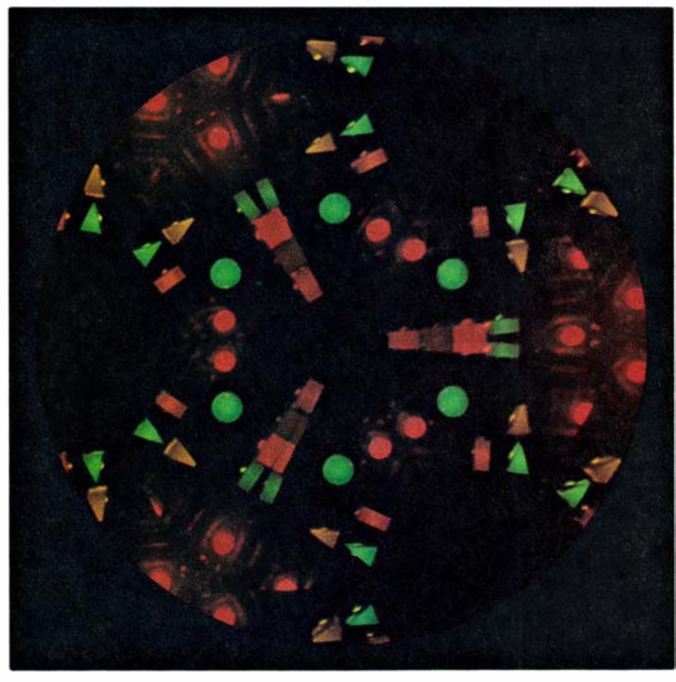
1. Imagen de calidoscopio de 16 sectores



2. Efecto de una lente



3. Objetos inmersos en aceite transparente



4. Imagen producida con diodos que reaccionan ante el sonido

En el suelo, y entre ambos espejos, deposité una pequeña caja y luego conté sus imágenes. El número de éstas aumentaba conforme disminuía el ángulo. Para un valor de 60 grados podían verse cinco imágenes. Se veían, además, imágenes de los cantos de los espejos, dos en cada lado.

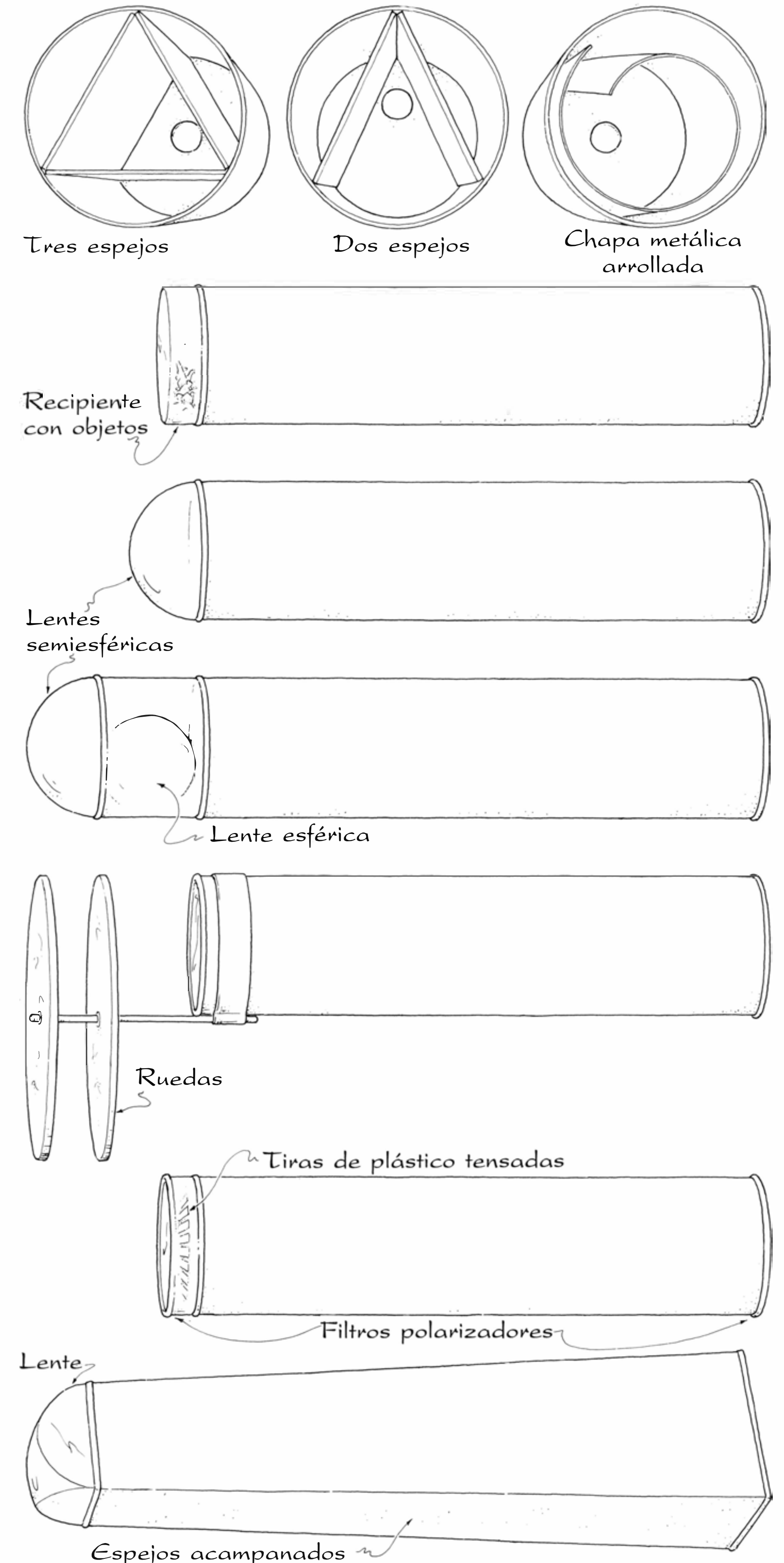
Seguidamente contemplé la caja colocándome por encima de los espejos. De nuevo volví a ver cinco imágenes además de la visión directa. Entonces comencé a entender de qué modo funciona un calidoscopio de dos espejos. Cuando las superficies reflectoras forman un ángulo de 60 grados, el instrumento proporciona seis vistas del objeto situado en el extremo opuesto. Estas vistas yacen en sectores en forma de porciones de tarta distribuidos en torno a un eje o centro, coincidente con la intersección de los espejos.

De estas vistas, una es directa y las otras son imágenes de reflexión. Pero todas aparentan encontrarse en el plano definido por la visión directa –el suelo, en este caso– al que llamaré plano de imagen. En los calidoscopios de juguete las superficies reflectoras son de mala calidad y las imágenes se presentan apagadas y borrosas comparadas con la visión directa. Mis espejos daban unas imágenes casi tan luminosas y nítidas como la visión directa.

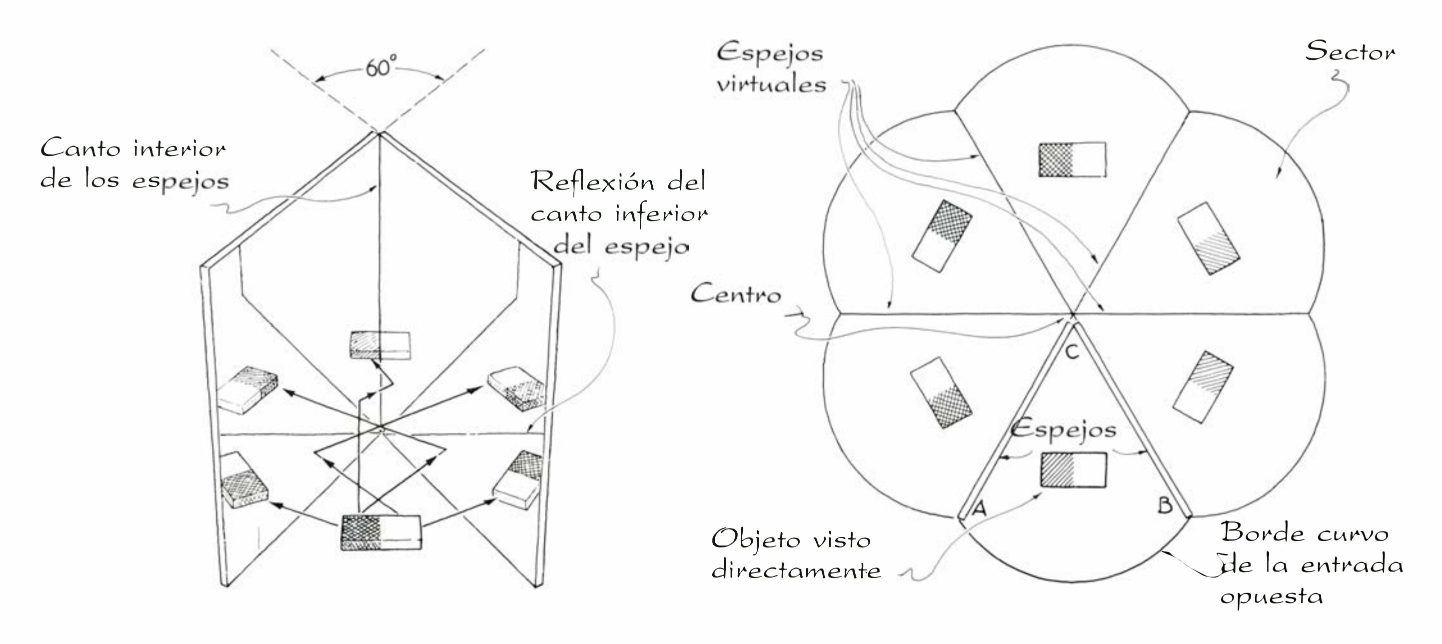
Al disminuir el ángulo entre los espejos, los sectores se contraían a medida que aparecían más imágenes en el lado del eje opuesto al de la visión directa. Los calidoscopios de juguete acostumbran a construirse para que no den más de cinco imágenes de reflexión, ya que las adicionales resultan excesivamente apagadas y borrosas. Pero con mis espejos logré observar imágenes nítidas aun cuando el ángulo valiese sólo 22,5 grados, con lo que conseguía una visión directa y 15 imágenes de reflexión.

¿Cuántas imágenes aparecen para un ángulo dado? Con dos espejos se puede dilucidar la respuesta. (Hay textos de física que contestan erróneamente. Contrariamente a lo que afirman, el número de imágenes no es necesariamente una menos que el cociente de 360 grados dividido por el ángulo entre los espejos.)

Un procedimiento sencillo consiste en trazar un esquema de lo que se ve, a partir de la visión directa. El centro o vértice es el eje del conjunto completo de imágenes. La visión directa se refleja en uno de sus espejos para formar un nuevo sector contiguo. Esta reflexión equivale a rotar la visión directa alrededor del espejo hasta que yaga



5. Diversos diseños de calidoscopios



6. Sistema de dos espejos

7. Agrupación producida por dos espejos

nuevamente en el plano del campo imagen. Hacia el otro lado de la visión directa se hace otra reflexión. Se prosigue reflejando cada nuevo sector en torno a su lado más apartado, a la vez que nos movemos en uno y otro sentido alrededor del eje, hasta que los sectores se encuentren o se solapen. Los límites radiales de cada sector, que en realidad son las imágenes de reflexión de los cantos inferiores de los espejos, se llaman espejos virtuales porque dan la impresión de reflejar cada sector en el siguiente.

Si al final los sectores se encuentran pero no se solapan, el número de imágenes es fácil de calcular: habrá una por sector. Pero si se solapan, el cálculo es más difícil. Consúltese el estudio de An-Ti Chai, de la Universidad Politécnica de Michigan, que se cita en "Bibliografía".

Continué con mis investigaciones añadiendo un tercer espejo a los dos primeros de modo que formaran un triángulo equilátero. Pensé que el montaje daría dos agrupaciones más de seis imágenes cada una, ya que había dos puntos más (*A* y *B*) que podían servir de centros. Cuando miré dentro del sistema, me encontré con el plano de imagen completamente cubierto de agrupaciones de seis imágenes. El espectáculo me dejó tan estupefacto que me esforcé en introducir la cabeza dentro del montaje todo lo que pude. Se distinguían al menos 20 agrupaciones que se extendían en todas direcciones dentro del plano de imagen, haciéndose cada vez más pequeñas en razón a su distancia a la visión directa [*véase la figura 8*]. Aún más hacia afuera pude ver docenas de imágenes que resultaban demasiado pequeñas para distinguirlas con claridad.

El campo imagen se componía de triángulos cuyos contornos correspondían a las reflexiones de los bordes inferiores de los espejos. Dentro de los triángulos había imágenes de la caja. Superpuestas al campo se veían líneas correspondientes a las imágenes de las intersecciones verticales interiores de los espejos. En todo el campo imagen se reproducían las agrupaciones alrededor de los puntos *A*, *B* o *C* de la visión directa. El campo completo podía construirse comenzando con la visión directa, reflejándola repetidamente respecto a cada uno de sus límites hasta llenar el plano.

En aquel momento yo no entendía cómo se reflejaban los rayos de luz en los tres espejos hasta llenar el plano de imagen. Me daba cuenta, empero, de que la óptica era la misma, en general, que la necesaria para formar las imágenes en un sistema similar de dos espejos paralelos [*véase la figura 9*].

Consideremos los rayos que parten de la cajita situada en el extremo opuesto del sistema de dos espejos paralelos. Los que van directamente al observador producen la visión directa. Otros parten de la caja bajo distintos ángulos con respecto al eje central del sistema. Fijémonos en el rayo que se refleja en el punto medio de uno de los espejos. Al ser interceptado por el observador, este rayo se percibe cual si tuviera su origen en un punto que resulta de extrapolar hacia atrás el rayo que le entra en el ojo. Ese punto se halla en la intersección de la recta extrapolada con el plano de imagen; la imagen se percibe como si estuviera allí.

Habrá otro rayo que salga de la caja y se refleje primero en el espejo de abajo y luego en el de arriba, en ambos casos a un cuarto de la longitud del espejo desde un extremo. El observador volverá a creer que el rayo procede del plano de imagen, merced a una extrapolación hacia atrás del rayo que le entra en el ojo. Pero ahora la posición estará más alejada de la visión directa que en el caso de un rayo que se refleje sólo una vez.

En la ilustración puede verse otro rayo que se refleja tres veces en los espejos. Primero en el espejo de arriba, a una distancia del extremo opuesto equivalente a la sexta parte de la longitud total. Después, en el punto medio del espejo de abajo. Y, por último, en el espejo de arriba, en el punto separado del extremo próximo una sexta parte de la longitud total. Tal rayo creará una imagen aún más alejada de la posición verdadera de la caja. Es así como se van creando imágenes ascendiendo por el plano de imagen. Cada imagen requiere tantas más reflexiones cuanto más separada está de la visión directa. En teoría no hay límite al número de imágenes; en la práctica, un número suficiente de reflexiones acaba emborronándolas. Descendiendo por el plano de imagen se creará otro conjunto similar de imágenes. En los calidoscopios de dos pares de espejos paralelos (menos corrientes) las imágenes se extienden también a derecha e izquierda del plano de imagen.

Sin duda, un sistema de tres espejos debe trabajar de una manera parecida; pero, ¿cómo, exactamente? Me decidí a seguir el trayecto de los rayos de luz desde que penetraban en el sistema

hasta llegar a mis ojos, reflejándose en los espejos. Por un momento consideré la posibilidad de colocar un láser en un lugar lejano para estudiar las reflexiones del haz al variar el ángulo de entrada, pero esta idea la abandoné ya que no era mi deseo que el haz pudiera irme a parar al ojo. Con las luces de la habitación apagadas, empecé a tantear el sistema con un láser de poca potencia apuntado hacia abajo.

Sabía yo que el trayecto de un rayo de luz desde la caja hasta mi ojo (incluso con reflexiones intermedias) era el mismo que desde mi ojo hasta la caja. Así que me apreté el costado del láser (de poco peso) contra la sien, cerré el ojo opuesto y moví la cabeza hasta apuntar el haz a una imagen de la caja. El haz atravesó el sistema de espejos, iluminando la caja y, por ello, añadiendo un punto rojo a todas las imágenes.

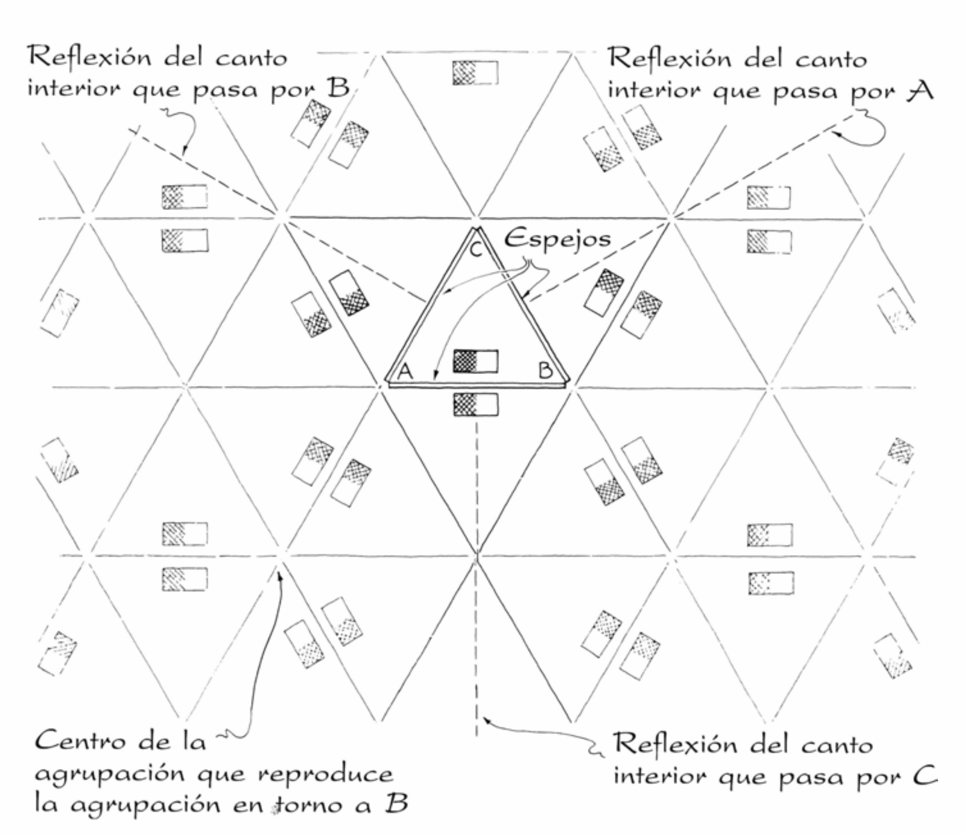
El dispositivo no era perfecto, pues el aparato me dejaba el costado de la cara algo desplazado con relación al trayecto de los rayos de luz que viajaban desde cada imagen concreta hasta mis ojos. Sin embargo, aquel pequeño ángulo de error careció, en general, de importancia, salvo que el haz se reflejara 10 o más veces. Para hacer visible el trayecto seguido por el haz introduje humo dentro del sistema.

Examiné primero la agrupación contigua a la visión directa, apuntando sistemáticamente el láser a cada imagen de la caja en la agrupación y observando, entonces, de qué modo el haz se entrecruzaba por los espejos hasta alcanzar la caja. Tracé un croquis de los trayectos e invertí luego las flechas que me indicaban el sentido de propagación de la luz [*véase la figura 10*]. Conseguí así un esquema de la manera habitual en que los rayos abandonan la caja y se reflejan en los espejos para dar lugar a los sectores de las agrupaciones. Así, el sector *J*, contiguo al de la visión directa, lo producen rayos de luz que se reflejan en el extremo opuesto del espejo izquierdo. El sector *K* se debe a los rayos reflejados en el extremo más alejado del espejo de abajo y, después, en el espejo izquierdo. El sector *O* se produce de forma similar, pero con la primera reflexión en el extremo más alejado del espejo derecho.

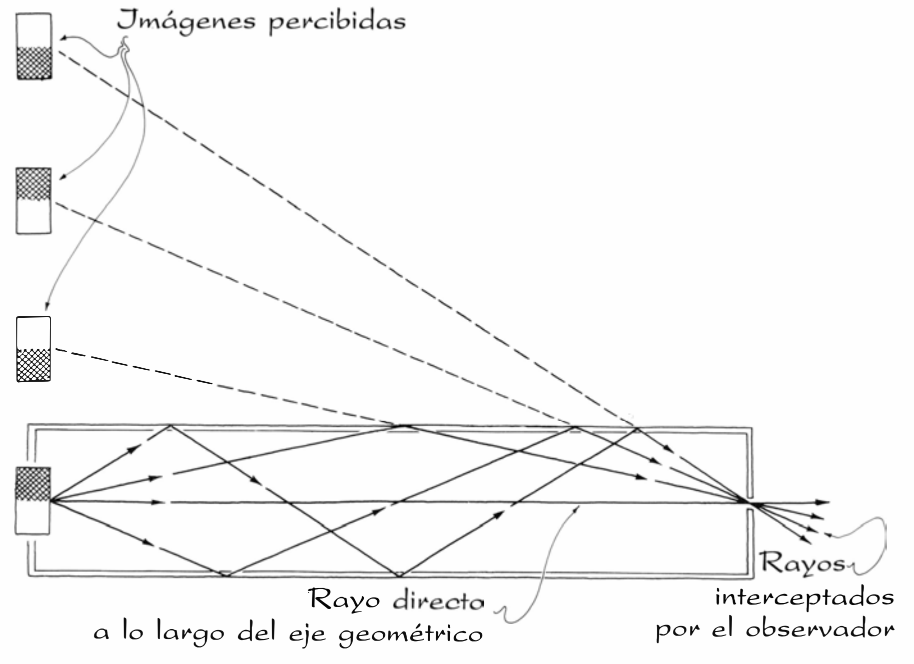
Los sectores *L* y *N* requieren tres reflexiones. Para el sector *L* los rayos se reflejan en el espejo derecho, luego en el de abajo y, finalmente, en el izquierdo. Para el sector *N* se reflejan primero en el espejo de abajo, después en el derecho y por último en el izquierdo. El sector *M* es más complicado, porque los rayos se reflejan cuatro veces: en el espejo de abajo, en el derecho, otra vez en el de abajo y por fin en el izquierdo.

Cuando hube confeccionado esa ilustración comencé a apreciar por qué los calidoscopios de tres espejos producen figuras más abigarradas que los de dos espejos. Si se retira el espejo de abajo, sólo quedarán las secciones *J* y *O*, puesto que son las únicas independientes de las reflexiones en ese espejo.

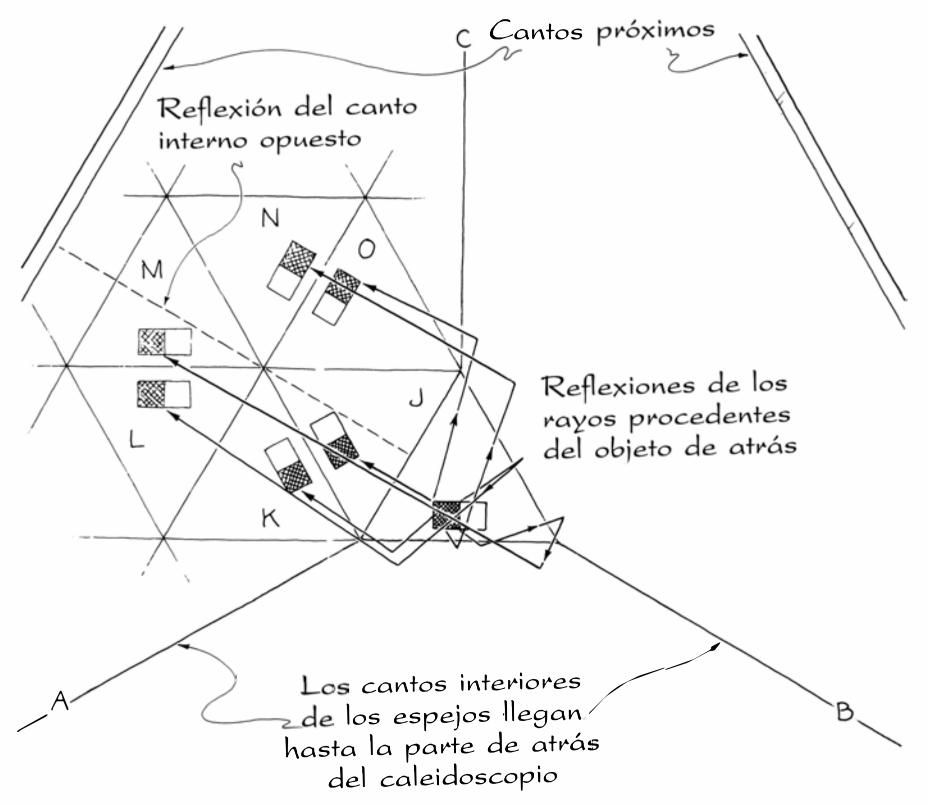
Proseguí explorando el campo imagen con mi haz de láser, contando el número de reflexiones asociadas a cada sector [*véase la figura 11*]. Tal como esperaba, a partir de mi estudio de un sistema de espejos paralelos, los sectores más separados de la visión directa requieren más reflexiones. Observé, asimismo, una curiosa secuencia en las reflexiones. Elijamos un centro cualquiera y consideremos las reflexiones que necesitan los sectores que lo ro-



8. Imágenes formadas por espejos dispuestos en triángulo equilátero



9. Reflexiones múltiples en espejos paralelos

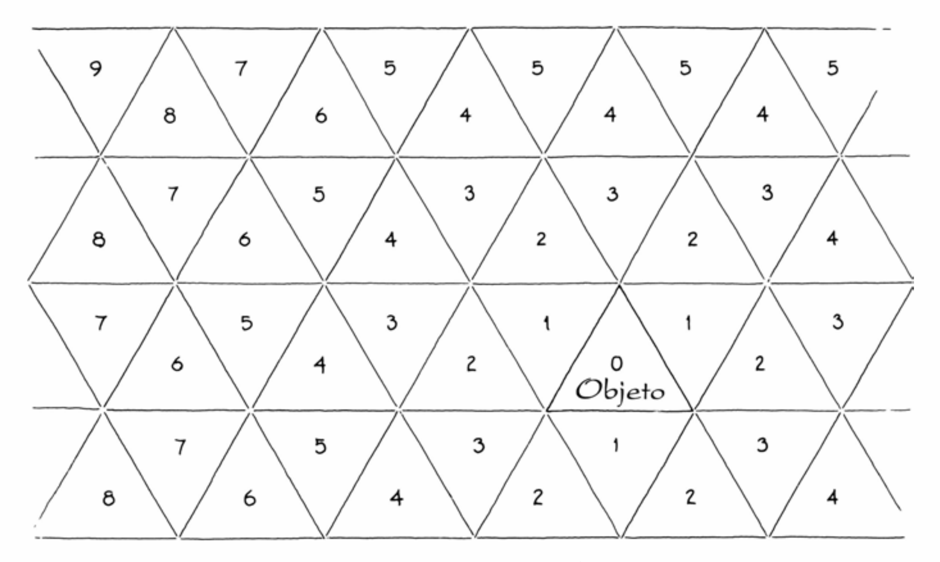


10. Reflexión de los rayos luminosos en un montaje en triángulo equilátero

dean. Tal como era de esperar, el sector más próximo a la vista directa requiere el número de reflexiones más bajo (lo llamo n) de esa agrupación concreta.

Movámonos ahora alrededor del centro hasta el siguiente sector, tanto en sentido horario como antihorario. Como estos sectores están más separados de la visión directa, requieren una reflexión más: $n+1$. Movámonos ahora un sector más en torno al eje en sentidos opuestos. Estos sectores requieren otra reflexión: $n+2$. Movámonos, finalmente, al sector más alejado de la visión directa. Este requiere una reflexión adicional: $n+3$. Resulta que todas las agrupaciones poseen una secuencia matemática como ésta alrededor de su centro.

Cuando la visión directa se compone exclusivamente de objetos que yacen en el plano de los extremos finales de los espejos, todos los sectores en el campo imagen son iguales. Si los objetos están más alejados, el contenido de los sectores debe ser distinto. De esta variación me había dado cuenta cuando saqué al aire libre mi sistema de espejos, lo puse horizontal sobre una mesa y lo apunté a la calle. Todos los sectores diferían en su contenido.



11. Número de reflexiones que crean los distintos sectores

Esta variación resulta de la perspectiva de la escena externa que se presenta en el extremo opuesto del sistema de espejos. En la figura 13 se muestra un ejemplo sencillo de esto. Dos objetos, representados por un círculo y un rectángulo, se hallan en el extremo del sistema más alejado del observador. El círculo no puede verse en la visión directa porque está demasiado desplazado del eje geométrico del sistema. Puede, no obstante, enviar rayos al interior del sistema bajo ángulos agudos con el eje geométrico. Estos rayos se reflejan tres veces antes de llegar al observador, quien percibirá el círculo cual si éste yaciera en un sector del plano de imagen distante de la visión directa. Análogamente, otros objetos alejados del eje geométrico mandarán sus rayos al interior del sistema bajo ángulos distintos y se percibirán como si estuvieran en otros sectores.

Los espejos abocinados del calidoscopio de Grannis y Lazarowski deforman el campo imagen en algo que tiene la apariencia de una esfera geodésica tridimensional. Para examinar esta deformación me puse en pie ante uno de mis espejos mientras lo inclinaba separando de mí el canto superior. Entonces, la imagen del suelo se inclinó dando la impresión de que el canto inferior del espejo se hallaba al borde de un precipicio.

Análogamente, en el calidoscopio de espejos acampanados cada sector aparenta estar inclinado separándose de mí, poque la forma de los sectores se presenta como un triángulo equilátero visto con cierta inclinación. Cuando desplazaba la vista hacia un lado de la visión directa los sectores parecían inclinarse más, creando la ilusión de hallarse sobre el costado de la esfera geodésica. Los rayos que se extienden radialmente hacia fuera desde la esfera son en realidad las imágenes de las líneas donde intersecan los cantos de los espejos. El fondo luminoso e indistinto que rodea a la esfera es la reflexión de unos espejos en otros.

Examiné también los campos imagen dados por espejos sin acampanar dispuestos en triángulo isósceles. Con dos espejos formé los dos lados iguales del triángulo y un tercero me sirvió de base. Al mirar dentro del sistema disminuí el ángulo entre los lados. De este modo, en la agrupación primaria alrededor de la intersección de los lados, aparecieron más imágenes; el resto del campo imagen se llenó de reflexiones incompletas de la escena en total desorden.

Para un ángulo de 22,5 grados, la agrupación primaria se componía de la

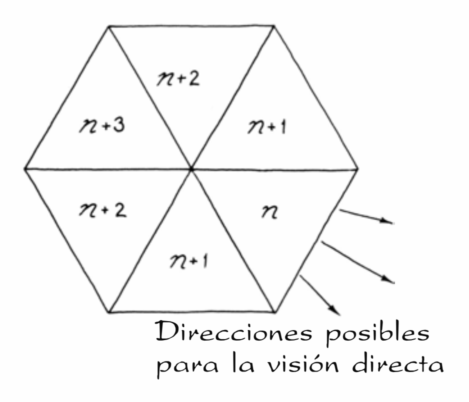
visión directa y 15 reproducciones de la misma (tanto exactas como reflejadas). El resto del campo imagen se componía sólo de porciones de esa agrupación [*véase la figura 14*]. Al mirar al espejo de la base vi el centro de la agrupación primaria reflejada en él. Modificando la visual cambiaba la porción visible de la agrupación primaria.

El ángulo de visión determinaba lo que podía verse en otras zonas del campo. Una manera de explicar el resto del campo es recurriendo al concepto de espejo virtual. Obsérvese que el espejo de la base está reproducido en todos los sectores de reflexión de la agrupación primaria. Cada una de estas reproducciones del espejo de la base hace de espejo virtual, capaz de reflejar la agrupación primaria. (Esta reflexión equivale a rotar la agrupación primaria en torno al espejo virtual hasta que vuelva a encontrarse en el plano de imagen.)

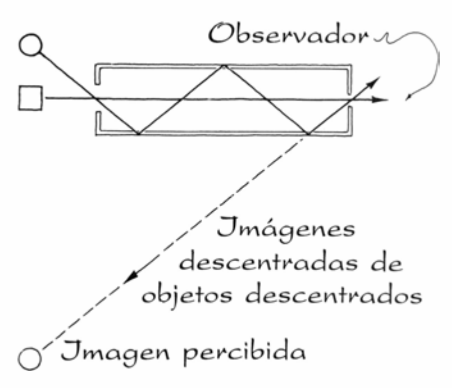
Al mirar directamente a un espejo virtual cual si éste fuera real, podía ver el centro de la agrupación primaria, pero espejos virtuales contiguos eliminaban el resto de la reflexión de ese espejo. Donde debería hallarse el resto de la agrupación se veían otras perspectivas de ésta reflejadas por los espejos virtuales contiguos. (Esta dependencia de la perspectiva no se indica en la ilustración.) El más caro de mis calidoscopios produce figuras así. Al principio creí que debía tener 16 espejos. (Para comprobarlo no me atreví a abrir un instrumento de su precio.) Pero ahora sé que tiene sólo tres espejos fuertemente reflectores.

Este tipo de campo imagen es fundamentalmente distinto del que forma un triángulo equilátero. Se trata de un campo ambiguo, en el sentido de que el contenido en un lugar concreto del campo depende de la inclinación de la visual bajo la que se observa el sistema. Un calidoscopio de triángulo equilátero produce un campo imagen no ambiguo, porque su contenido en un sitio concreto no varía cuando cambia la inclinación de la visual.

Los campos imagen ambiguos no me gustan tanto, pese a la bella simetría de la agrupación primaria. ¿Hay más configuraciones de tres espejos que den campos nítidos? ¿Pueden generarse éstos con sistemas de más de tres espejos? A los lectores dejo esta pregunta acompañada de otra que les ayudará: ¿Existe una figura geométrica que pueda reflejarse repetidamente, en cada uno de sus lados, hasta llenar un plano sin solaparse? El mes que viene les revelaré la solución general que he hallado.

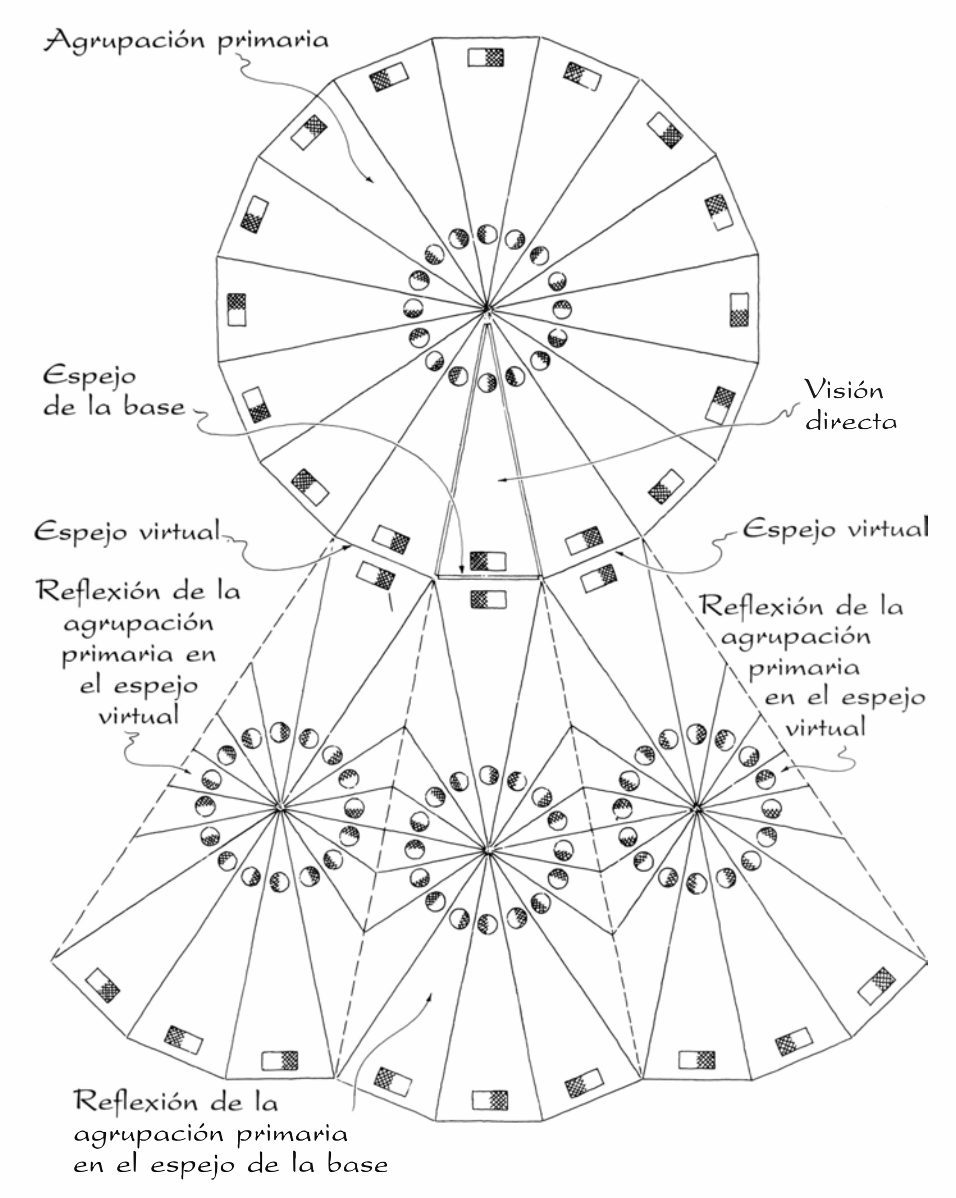


12. Secuencia de reflexiones en una agrupación



13. Formación de la imagen de un objeto situado ante un calidoscopio

También puede ser entretenido estudiar sistemas de espejos dispuestos en polígono regular de número par de lados. Aquí el campo imagen debe ser similar al generado por dos espejos paralelos. ¿Existen diferencias? ¿Qué ocurre cuando el polígono tiene un número impar de lados? ¿Qué se verá en un calidoscopio cuyo interior sea un cilindro reflector, lo que supone un polígono de infinito número de lados? ¿Cabe esperar de un calidoscopio de espejos curvos que dé campos inambiguos?



14. Imágenes en un calidoscopio con los espejos en triángulo isósceles

Libros

Caos, etología de los póngidos, ciencia de la Ilustración y materiales magnéticos

Carlos Pérez, Andrés de Haro, Antonio Lafuente y M. San Miguel

UNIVERSALITY IN CHAOS, por Predrag Cvitanović. Adam Hilger Ltd.; Bristol, 1984. El diccionario de la Real Academia define el caos como aquel estado de confusión en que se hallaban las cosas al momento de su creación, antes de que Dios las colocase en el orden que tuvieron después. Y añade dos sinónimos: confusión y desorden. Una definición bastante imprecisa, que en física comienza a adquirir contornos más netos.

El desorden, la función desorden, tiene un significado claro en mecánica estadística: es el número de estados, en el nivel microscópico (microestados), compatibles con un estado dado en el nivel macroscópico (macroestado). Boltzmann postuló, en el siglo pasado, una fórmula que relaciona la entropía con el desorden para sistemas en equilibrio termodinámico. Así, el crecimiento de la entropía equivale a un aumento del desorden. La inversa de esta fórmula, utilizada por Einstein por vez primera, permite asociar una probabilidad a las fluctuaciones de las magnitudes que definen el estado de un sistema.

Este concepto –el desorden– y esas fórmulas son puramente estáticos, como lo es la termodinámica del equilibrio. Sólo sirven para estudiar y comparar estados de equilibrio o muy próximos a él. Para seguir la evolución de un sistema alejado del equilibrio hay que fijarse en su dinámina, en la cinética de los procesos que intervienen. Esa evolución no se rige únicamente por las fluctuaciones, sino fundamentalmente por la dinámica intrínseca del sistema. Según sea ésta se pueden dar inestabilidades, una organización más compleja o una turbulencia. Fijémonos, por ejemplo, en lo que pasa en una tubería. Cuando el caudal es pequeño, el movimiento del líquido en ella es bastante ordenado: el flujo es laminar; al aumentarlo, llega un momento en que el flujo se torna turbulento: ha tenido lugar una transición. En la turbulencia, el comportamiento de un elemento del fluido es impredictible, y para describirlo es necesario aumentar el número de variables.

Hasta aquí se han descrito dos categorías de movimientos desordenados: las fluctuaciones estocásticas y la turbulencia. Pues bien, el caos es una nueva categoría, con características similares, aunque con otras diametralmente opuestas. Ciertos sistemas disipativos pueden describirse mediante unas pocas ecuaciones deterministas –pocos grados de libertad, a diferencia tanto del desorden como de la turbulencia–, es decir, sin incluir en la descripción elementos estocásticos, y cuyas soluciones presentan extrañas formas, para ciertos valores de los parámetros. Lo fundamental de esas soluciones es una curiosa sensibilidad a las condiciones iniciales: cambiándolas ligeramente, la nueva trayectoria sigue siendo parecida en la forma, pero, al poco, la correlación entre ambas se pierde. Esas trayectorias reciben el nombre de atractores extraños o atractores fractales, y su singular comportamiento se denomina caótico o, simplemente, caos. Mientras la turbulencia es impredictible a corto plazo, el caos manifiesta una impredictibilidad transcurrido cierto tiempo.

Entre los científicos está fuertemente arraigado el determinismo: conocidas las ecuaciones de evolución de un sistema y las condiciones iniciales, podemos saber cuál será su estado en cualquier otro instante. Poincaré, a finales del siglo pasado, advirtió que esta afirmación sólo sería exacta si se pudiera conocer con precisión absoluta, sin indeterminación alguna, las condiciones iniciales. Tras el advenimiento de la mecánica cuántica y, más en concreto, del principio de incertidumbre de Heisenberg, sabemos que eso no es posible. En un sistema con comportamiento caótico, la más mínima indeterminación inicial crece exponencialmente en el tiempo.

Existe una gran variedad de sistemas en los que se ha hallado comportamiento caótico: ciertos tipos de láseres, circuitos electrónicos, sistemas fluidos, las células del corazón y ciertas reacciones químicas, por sólo citar unos cuantos ejemplos. En esos sistemas se trata ahora de discernir entre las características comunes, universales, y aquellas otras que son singulares. El intento es similar a la clasificación de los puntos críticos en clases de universalidad llevada a cabo en los años setenta. Los métodos de clasificación del caos son todavía materia de discusión.

El libro que nos ocupa pretende allanar el camino a quienes se sientan atraídas por tan apasionante tema. Es un compendio de los cuarenta y un artículos que, a juicio del editor Predrag Cvitanović, han marcado un hito en la comprensión de algún aspecto del caos. Los conocedores del tema estarán de acuerdo –salvo en pequeños detalles– con la selección hecha por Cvitanović, ya que ha sabido recoger los artículos más claros e interesantes, muy dispersos en la bibliografía. Quizá los matemáticos echen de menos contribuciones de algunos de sus colegas, en particular trabajos de Smale, de Guckenheimer o de Arnold; como el editor advierte en el prólogo, el libro no es exhaustivo.

La introducción, del propio editor, es insuficiente, por parcial, limitándose a comentar un modelo determinado. Le falta visión de conjunto y no comenta los interesantes aspectos filosóficos que el nuevo concepto entraña, en especial con relación al determinismo materialista.

Los artículos están clasificados en siete partes, más una octava dedicada a referencias. Aunque respeto el criterio del autor, quizá hubiese sido preferible situar el artículo de N. Lorenz (recogido en la página 292) justo al comienzo, ya que se trata del trabajo que ha originado los estudios sobre el caos. Están bien elegidos, sin embargo, los artículos de la primera parte, en especial el de Ruelle sobre los atractores extraños.

En la segunda parte se agrupan los artículos experimentales, comenzando por los de la mecánica de fluidos: inestabilidades de Rayleigh-Bénard y de Taylor-Couette. Se echa en falta alguno de los magníficos artículos de revisión del problema de Bergé y Dubois.

La segunda parte consta de cuatro secciones. Se encuentran en ellas los artículos más relevantes sobre el caos en sistemas químicos, en especial en la reacción de Belusov-Zabotinski, en experimentos ópticos –falta algún artículo sobre el caos en biestabilidad óptica–, experimentos en circuitos electrónicos y, por último, en sistemas biológicos.

La tercera parte está formada por artículos teóricos relacionados con la universalidad y el desdoblamiento de pe-

ríodos. Como nota curiosa cabe reseñar la anécdota que se recoge al final de un artículo de Feigenbaum, el mejor que tiene sobre el tema: lo finalizó en 1976 y, tras ser rechazado en dos revistas de matemática, se publicó en el *Journal of Statistical Physics* en 1979. En una edición posterior habrá que agregar las importantes contribuciones de Farmer y Procaccia sobre la dimensión fractal de los atractores extraños.

En la cuarta se encuentran artículos sobre ruido: ruido determinista –aunque parezca paradójico– y efectos del ruido externo sobre las diversas inestabilidades. En esta sección hubiera sido conveniente incluir algún trabajo experimental.

Dos artículos referentes a la transición hacia la turbulencia por intermitencia constituyen la parte quinta. Se resiente de la ausencia de artículos experimentales, particularmente los de Dubois sobre el problema de Bénard en pequeños recipientes. Las partes sexta y séptima son una excusa para incluir algunos artículos interesantes. La última parte, de recopilación bibliográfica, pone de nuevo de manifiesto las preferencias del autor por el modelo de Feigenbaum.

Se trata, en suma, de un material bibliográfico muy útil para investigadores que quieran trabajar sobre comportamientos caóticos: físicos, químicos, ingenieros, matemáticos, etcétera. Los artículos de la primera parte resultan asequibles a los principiantes, pero a partir de la tercera comienzan a ser mucho más especializados. El editor añade pequeños comentarios al final de algunos artículos, que ayudan a situarlos en el contexto. La presentación está cuidada.

El criterio de ordenación y de selección es bastante parcial. Tanto en la introducción como en algunos artículos se nota que el editor ha sido discípulo de Feigenbaum, cuya clasificación del comportamiento caótico, aunque importante, no es, ni con mucho, definitiva. A pesar del esfuerzo realizado, el comportamiento caótico sigue resistiéndose al orden clasificador. (C. P.)

ETOLOGÍA DE LA VIVIENDA HUMANA, por Jordi Sabater Pi. Editorial Labor; Barcelona, 1985. Este libro representa una preocupación por integrar el significado etológico de la construcción de nidos para descansar realizados por los póngidos y la presencia de construcciones habitadas por los hombres primitivos, con una mentalidad evolutiva y filogenética, propia del método etológico. El autor se pregunta sobre las relaciones conductuales de los nidos de los póngidos, ubicados en espacios tróficos, de los puntos fijos y hogares-base que perduran en los rompevientos y de los simples habitáculos de los hombres primitivos fósiles y actuales.

La documentación es abundante y de primera mano, habiendo estudiado 448 camas-nido de gorilas en Río Muni y 195 nidos de chimpancés en el mismo enclave. Expone, hasta los pormenores, la estructura y técnicas seguidas en su elaboración. En los gorilas los nidos son ovalados y se construyen en áreas llanas recubiertas de formaciones de *Aframomum*, en áreas arbustivas o donde crece *Musanga cecropioides*, árbol típico del bosque agrológico. Describe las técnicas de construción y destaca el aspecto de la mullidez, confortabilidad y aislamiento, especialmente los ubicados en el suelo. Los nidos de los chimpancés son también ovalados, con estructuras más complejas que las de los gorilas, por tratarse de plataformas instaladas casi siempre en los árboles y a bastante altura.

En el estudio realizado sobre los materiales empleados en la construcción de sus camas, destaca la mayor utilización de *Aframomum* por parte del gorila, que se justifica por ser ésta una especie vegetal muy apta para ello debido a su suavidad, abundancia de hojas grandes durante todo el año, flexibilidad de los largos tallos que pueden ser doblados, cruzados o arrancados con facilidad, y también porque esta planta forma extensos macizos utilizados dietéticamente (comen sus hojas tiernas, médulas y frutas durante todo el año). La necesidad de nidificar se realiza al atardecer, en áreas donde los gorilas se alimentan intensamente a estas horas del día. No obstante, se considera que la bondad del material utilizado en la construcción prima sobre el condicionamiento dietético, por la comprobación de grupos de gorilas que al atardecer se alimentaban en el bosque denso y que se desplazaban a los campos de *Aframomum* para nidificar en ellos.

La determinación del material de nidificación en chimpancés es más dificultosa por la situación de sus nidos a gran altura en áreas de bosque denso o secundario. Tanto gorilas como chimpancés hacen sus camas al atardecer y las abandonan al amanecer. Ambas especies tardan unos 5 minutos en la construcción del nido. Es de destacar que ambos antropoides fabrican al mediodía camas para descansar durante las horas de inactividad; en cuyo caso se trata de estructuras muy elementales, simples cojines de vegetación.

Los gorilas buscan el beneficio de una cobertera en la estación húmeda. En el chimpancé también es muy significativa la construcción de nidos debajo de un cubierto formado por un estrato superior de la vegetación, que proporciona cierta protección o resguardo contra la lluvia y el rocío. En ambas especies no se da depedencia entre la orientación del nido y las estaciones climáticas. El gorila, y a veces el chimpancé, deposita los excrementos dentro del nido.

Relacionadas con una posible organización en su distribución, resultan interesantes las anotaciones del autor sobre la ordenación de las camas dentro de las áreas de anidaje. Como llama también la atención el comensalismo, de gorila y el chimpancé, que obtienen alimento de la presencia humana.

Merecen destacarse las comparaciones realizadas entre la actividad nidificadora de los póngidos y la actividad constructora de refugios de los hombres primitivos. Y valiosos son los datos ofrecidos sobre los campamentos humanos nómadas de los cazadores-recolectores, aunque éstos son grupos humanos muy evolucionados y poseedores de cultura. Las páginas dedicadas al hombre fósil son muy sugestivas, pues es en los restos fósiles donde se han de buscar las estructuras primitivas que servían de descanso y cobijo durante el proceso de hominización y que podrán relacionarse con los nidos de los póngidos. Es éste un campo en el que está todo por decir: aun han de encontrarse construcciones que nos manifiesten una etología fosilizada y que nos digan cómo eran las construcciones de los primitivos homínidos y sus relaciones con los nidos de los póngidos.

El libro aporta numerosos datos objetivos de primera mano y ofrece sugerencias para rellenar ese hiato que va desde la construcción diaria de nidos para descansar por parte de los póngidos hasta la elaboración de refugios permanentes para ellos y para los alimentos por parte de los humanos. Representa un esfuerzo por aclarar los aspectos problemáticos de la etología de la conducta constructora de póngidos y humanos, al tiempo que nos muestra cuánto hay que ahondar en el significado de la gran diferencia motivacional que existe entre nido para descansar y vivienda para habitar. (A. de H.)

EPOCA, GÉNESIS Y TEXTO DE LAS "NOTICIAS SECRETAS DE AMÉRICA" DE LOS TENIENTES DE NAVÍO JORGE JUAN Y ANTONIO DE ULLOA, por Luis J. Ramos Gómez. Madrid, Consejo Superior de Investigaciones Científicas, Colección Tierra nueva e Cielo Nuevo, 1985. Sin

duda, las *Noticias secretas de América* redactadas por Jorge Juan y Antonio de Ulloa en 1747 siguen siendo una fuente primordial para el conocimiento de la realidad colonial americana durante el siglo XVIII. La obra, cuya circulación reservada por los distintos gabinetes ministeriales del setecientos español fue muy estimable, permaneció inédita hasta su impresión en Londres en 1826. Se trataba de un incisivo y documentado informe acerca del estado de la administración civil, militar, eclesiástica y económica, cuyos tonos críticos aconsejaban una prudente difusión interna y casi exclusivamente gubernamental. De ahí que cuando David Barry lograra el manuscrito en circunstancias poco conocidas, aplicase a las informaciones que contenía el calificativo de *secretas*, buscando una notoriedad pública que tal vez no hubiese logrado con un título más acorde al texto y que, según su actual editor Luis J. Ramos García, debiera haber sido *Discurso y reflexiones políticas sobre el estado presente de los reinos del Perú en lo tocante a gobierno, régimen particular de aquellos habitadores y marina.*

¿Quiénes eran Jorge Juan y Antonio de Ulloa y cuáles las razones de su estancia en América? Durante la primera mitad del setecientos, la ciencia europea se vio involucrada en una prolongada polémica que enfrentaba a newtonianos y cartesianos sobre el tema de la figura de la Tierra. Mientras en Inglaterra se afirmaba, de acuerdo con las teorías de Newton, que nuestro planeta estaba achatado por los polos, en Francia se sostenía con ardor la tesis de una Tierra oblonga. Lo que inicialmente comenzó siendo un debate académico, pronto habría de transformarse en un conflicto entre ciencias nacionales. A comienzos de los años treinta los acontecimientos se precipitan; en Francia, tras el repliegue de la Académie des Sciences sobre la ortodoxia neocartesiana, se decide organizar dos expediciones científicas a Laponia y Quito respectivamente, cuyos resultados tendrían que zanjar la polémica y decidir entre las dos cosmografías rivales.

Para poder realizar la expedición americana se necesitaba un permiso concedido por el Consejo de Indias que autorizara la libre circulación de los miembros de esta singular embajada europea en los dominios coloniales. Como consecuencia de las gestiones diplomáticas iniciadas por Francia en 1734, España impone la presencia de dos jóvenes cadetes guardamarinas de la Academia que fundara José Patiño en Cádiz en 1717. La misión que se les asigna bajo el término *asistir* a las operaciones de los académicos, consistía tanto en vigilar sus movimientos como en aprender y colaborar en el desarrollo de los trabajos geodésicos. En términos generales, el programa de observaciones que se iba a realizar involucraba a todos los métodos y prácticas de la geografía del momento; era, pues, muy apetecible para las autoridades de marina la participación de dos españoles que pudieran efectuar un aprendizaje que ninguna institución de las aquí existentes podía asegurar. Pero ello no es todo; paralelamente se les encomienda en uno de los diez puntos de la *Instrucción* reservada que firmó Patiño en abril de 1735 lo siguiente: "Levantarán planos de las ciudades y puertos, con sus fortificaciones, donde hicieren asiento, y se informarán de los términos de sus provincias y gobernación, de los pueblos o lugares que contiene, y de lo fértil o estéril de sus campos, como también de la inclinación, industria y habilidad de sus naturales, y la braveza o joviendad de los indios irreductos, y facilidad o dificultad de su reducción". La orden, sin duda, constituye una buena muestra acerca de las preocupaciones que poseían los responsables de la gobernación colonial.

Quedaban, en consecuencia, nuestros marinos, autorizados para recorrer la colonia sin vendas en los ojos; más allá de la rutina administrativa o del mero informe técnico, quería el secretario de Indias un juicio que no estuviese mediatizado por los compromisos y corruptelas de las autoridades locales. Es difícil saber qué esperaba Patiño allá por 1735 de esta comisión a dos cadetes, precipitadamente ascendidos a tenientes de navío, que sólo contaban con 21 y 19 años respectivamente.

Sin embargo, no habrían de resultar un simple trámite las instrucciones que se les dieron. De su participación en los trabajos de la expedición y otras comisiones que les encomendara el virrey de Lima, surgieron varias obras en las que se interpenetran los dos objetivos fundamentales de su viaje. Las *Observaciones astronómicas y físicas*... y la *Relación histórica del viaje a la América meridional*..., aparecidas en 1748, configuran uno de los polos de su trabajo. Además de todas las operaciones de carácter científico, nos proporcionan las más preciosas noticias de carácter antropológico, sanitario, etnográfico, urbanístico y geográfico de la mitad sur del subcontinente americano. En el otro polo, además de las ya citadas *Noticias secretas*..., publicarán una obrita en la que, bajo el título *Disertación histórica y geográfica sobre el meridiano de demarcación* (1749), se ocuparán de uno de los temas más acuciantes y conflictivos de la política internacional de España. Ciencia y política se entrelazaban sin violencia. Jorge Juan y Antonio de Ulloa irrumpían en la escena cultural española encarnando un nuevo modelo de ilustración, hasta entonces inédito entre nosotros.

¿Cómo fue posible? La vida de los expedicionarios en la Real Audiencia de Quito no iba a ser fácil. Nada más llegar a la capital fueron acusados de introducir mercaderías ilegalmente; Couplet, uno de los franceses enviados, muere de fiebres malignas; su dotación de recursos financieros, mal calculados, se ve drásticamente reducida a un puñado de pesos que le obligarán a endeudarse. Seniergues, cirujano de la expedición, es asesinado en 1739 en Cuenca durante un motín en el que el pueblo enfervorecido gritaba "Viva el Rey, y mueran los Gavachos y otras voces". La guerra declarada en 1739 entre España e Inglaterra impide las comunicaciones con Europa y consolida una situación de aislamiento respecto de la Academia de París. La misión geodésica se convierte en una aventura que no discurre al margen de la realidad colonial. Los expedicionarios, y especialmente Juan y Ulloa, se ven obligados a integrarse no superficialmente en la vida colonial. Más aún, entre 1736 y 1740 una fuerte agitación política, religiosa, administrativa y bélica conmovió al país en el marco de una profunda crisis en las actividades económicas del virreinato. Los obrajes quiteños asistían a un proceso de decadencia que era atribuido a la ilegal competencia derivada de la sistemática introducción de mercaderías foráneas. Varias novedades habían avivado la vieja pugna entre las dos castas étnicas dominantes de los chapetones y criollos. La crisis del sector minero, agravada por el bajo rendimiento de Huencavélica, había producido un importante descenso en la demanda interna de productos manufacturados en América y, como consecuencia, se iniciaba un proceso de desintegración económica del virreinato que obligaba a las distintas regiones a buscar su integración en el mercado mundial. Esta regionalización de la economía tenía implicaciones que eran observadas con gran inquietud tanto en el interior del virreinato como desde la metrópoli. En fin, la guerra con Inglaterra probaba las graves deficiencias del sistema defensivo colonial y hacía más evidentes las contradicciones existentes, así como las injusticias y corruptelas de la administración. Este era en resumen el marco general en el que

se desarrollaron los trabajos de la compañía científica. Jorge Juan y Antonio de Ulloa supieron apreciar gran parte de lo que ocurría y tener el valor de escribirlo.

La edición que en dos volúmenes publica ahora el CSIC en su nueva colección *Tierra nueva e Cielo nuevo*, además de oportuna, es elogiable por muchas razones. La primera, porque no se ha desaprovechado una vez más la ocasión para hacerlo con seriedad y competencia. El primer volumen (440 pp.) contiene un documentado y muy afortunado estudio de Luis J. Ramos Gómez, en el que se analizan las circunstancias que rodearon la estancia de Juan y Ulloa en América. Su lectura es imprescindible para poder entender el contenido de las *Noticias secretas de América*. El segundo no se limita a reproducir el texto de uno de los manuscritos que aún se conservan; además de contrastar las distintas versiones, está plagado de jugosas notas a pie de página en las que se juzga el rigor o deformación con los que los dos marinos españoles describen la realidad.

Durante los años de estancia en América, algunos ilustrados quiteños pudieron beneficiarse del contacto con los académicos franceses. Más aún, sus actividades de carácter científico, cultural o técnico se vieron sólidamente respaldadas y legitimadas por el apoyo que pudo proporcionarles la prestigiosa institución parisiense. Sus estudios realizados en la soledad de algunos *parnasos* aislados y marginales y hasta entonces olvidados, adquirían una dimensión científica que los expedicionarios supieron reconocer. El estímulo de estos contactos fue suficiente para que emergieran en la colonia algunas importantes obras cuya objetivo genérico sería la identificación de un territorio y de un mercado. Destacan entre los autores Pedro V. Maldonado, Jean Magnin, Dionisio de Alsedo, Jacinto Morán Butrón e incluso Juan Pío Montufar. ¿No podría haberse contrastado el contenido de las *Noticias secretas de América* con estos escritos coetáneos? Tal vez, esta sería la única cuestión que se ha quedado pendiente para estudios posteriores. (A. L.)

THE THEORY OF MAGNETISM II. THERMODYNAMICS AND STATISTICAL MECHANICS, por D. C. Mattis. Springer Verlag; Berlín y Heidelberg, 1985. Esta obra constituye el volumen 55 de la Springer Series in Solid State Sciences, editada por M. Cardona, P. Fulde y H. J. Queisser. La primera parte, *The Theory of magnetism I. Statics and Dynamics*, del mismo autor, apareció como el volumen 17 de la misma serie. Esa primera parte trata de los aspectos microscópicos de la física de los materiales magnéticos y de los criterios para un comportamiento cooperativo macroscópico. En palabras del propio autor, la segunda parte, que aquí nos ocupa, considera los modelos más sencillos que permiten un análisis matemático a temperatura finita. Esta segunda parte es autónoma e independiente de la primera, aunque aparezcan referencias ocasionales a resultados examinados en aquélla. El libro es un tratado sobre modelos mecánico-estadísticos de transiciones de fase de tipo magnético, y de ahí su subtítulo. D. C. Mattis, arropado por una larga experiencia de investigación en este tema, hace una presentación coherente de resultados ya clásicos y simultáneamente introduce al lector en temas de interés actual. Se pretende familiarizar al lector con los métodos de la mecánica estadística. Esto se hace dentro del contexto de los sistemas magnéticos, no sólo por su propio interés, sino también por la aplicabilidad general de esos métodos.

El contenido de este libro importa a un buen número de estudiantes y profesionales por su valor en técnicas y aplicaciones de diversas áreas: matemática y física teórica, física del estado sólido, ciencia de materiales, química física, etcétera. Sin embargo, el libro no puede considerarse un texto elemental. Es indispensable un conocimiento básico de mecánica cuántica. Por lo que a la mecánica estadística se refiere, el libro pretende ser autosuficiente, introduciendo las nociones básicas en las dos primeras secciones. Aunque se consigue así un recetario útil y claro de cómo pasar de la descripción microscópica a la descripción termodinámica, parece que un conocimiento más profundo y sistemático de estas ideas debe permitir aprovechar mucho más los desarrollos de este libro. Un conocimiento básico de la mecánica estadística puede conseguirse, por ejemplo, mediante el volumen 30 de la misma colección (*Statistical Physics I*, por M. Toda, R. Kubo y N. Saito). De hecho, este otro libro trata de forma concisa y elemental, y dentro de un contexto más amplio, algún aspecto de los modelos considerados por Mattis. Desde este punto de vista, puede decirse que la obra de Mattis es un libro de consulta y referencia, apropiado para introducir a investigadores noveles en el tema y útil también para un tercer ciclo universitario.

El libro está dividido en tres capítulos. En el primero se hace un resumen, sección por sección, del contenido de la obra. El segundo capítulo introduce brevemente los rudimentos necesarios de la mecánica estadística y considera luego diversos aspectos de múltiples modelos de transiciones magnéticas. Incluye una concisa y clara explicación de la teoría de campo medio para sistemas ferromagnéticos y antiferromagnéticos, presenta los modelos gaussiano y esférico, expone la teoría de los vidrios de espín y considera los modelos XY clásico y cuántico con especial referencia a las consecuencias de la simetría continua en dos dimensiones. El tercer capítulo está esencialmente dedicado a resultados exactos para el modelo de Ising en una y dos dimensiones y a un resumen de resultados para el caso tridimensional. Esta estructura revela el carácter no elemental del libro al considerar el modelo sencillo de Ising después de modelos más complicados, en contra del procedimiento habitual. Puede sorprender, por ejemplo, que se introduzca el método de la matriz de transferencia para el modelo XY antes que para el modelo de Ising. Ello puede justificarse por el mayor nivel de complicación matemática utilizada en el tercer capítulo.

Un mérito importante de este libro es presentar un compendio de resultados y técnicas, que en general no son accesibles en forma de texto. Es de gran utilidad la inclusión de introducciones a temas de interés actual, como los vidrios de espín o las transiciones tipo Kosterlitz-Thouless en dos dimensiones, así como el uso repetido de métodos importantes como el de la matriz de transferencia o la representación fermiónica de los operadores de espín. En determinados temas avanzados o controvertidos, el autor se limita a describir resultados, pero dando en este caso una adecuada bibliografía. Como aspecto negativo podría quizá citarse la ausencia de algunos temas. Así, hay otros modelos, como los de Potts, ANNI, Z_N, etc., o fenómenos como metamagnetismo y estructuras espaciales que, cuando menos, merecerían recordarse. El propio autor justifica otras ausencias, como la del efecto Kondo, y comenta en el primer capítulo que el tema de los fenómenos críticos y grupo de renormalización, que no se ha abordado en esta edición, podría incluirse en una próxima.

Finalmente, cabe mencionar que el libro está jalonado de citas y consideraciones históricas que sitúan algunos temas en una perspectiva adecuada. También se incluye un buen número de ejercicios y problemas intercalados en el texto. (M. S. M.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

SOPORTE LOGICO PARA LA DEFENSA ANTIMISIL

SAFEGUARD DATA-PROCESSING SYSTEM. *The Bell System Technical Journal*, vol. 54, suplemento especial; 1975.

SOFTWARE ENGINEERING ECONOMICS. Barry W. Boehm. Prentice-Hall, Inc., 1981.

CONJUGACION DE FASE OPTICA

OPTICAL PHASE CONJUGATION. Dirigido por R. A. Fischer. Academic Press, 1983.

OPTICAL PHASE CONJUGATION. Número especial del *Journal of the Optical Society of America*, vol. 73, n.° 5; mayo, 1983.

PRINCIPLES OF PHASE CONJUGATION. B. Ya. Zel'dovich, N. F. Pilipetsky y V. V. Shkunov. Springer-Verlag, 1985.

COMUNICACION POR VIA ACUSTICA EN EL GRILLO

TYMPANAL MEMBRANE MOTION IS NECESSARY FOR HEARING IN CRICKETS. Hans-Ulrich Kleindienst, David W. Wohlers y Ole Naesbye Larsen en *Journal of Comparative Physiology*, vol. 151, págs. 397-400; 1983.

AUDITORY BEHAVIOR ON THE CRICKET, 3: TRACKING OF MALE CALLING SONG BY SURGICALLY AND DEVELOPMENTALLY ONE-EARED FEMALES, AND THE CURIOUS ROLE OF THE ANTERIOR TYMPANUM. Franz Huber, H.-U. Kleindienst, Theo Weber y John Thorson en *Journal of Comparative Physiology*, vol. 155, págs. 725-738; 1984.

TEMPORAL SELECTIVITY OF IDENTIFIED AUDITORY NEURONS IN THE CRICKET BRAIN. Klaus Schildberger en *Journal of Comparative Physiology*, vol. 155, págs. 171-185; 1984.

SIDA Y SISTEMA INMUNITARIO

IMMUNOREGULATORY LYMPHOKINES OF T HYBRIDOMAS FROM AIDS PATIENTS: CONSTITUTIVE AND INDUCIBLE SUPPRESSOR FACTORS. Jeffrey Laurence y Lloyd Mayer en *Science*, vol. 225, n.° 4657, págs. 66-69; 6 de julio de 1984.

LYMPHADENOPATHY-ASSOCIATED VIRAL ANTIBODY IN AIDS: IMMUNE CORRELATIONS AND DEFINITION OF A CARRIER STATE. Jeffrey Laurence, Françoise Brun-Vezinet, Steven E. Schutzer, Christine Rouzioux, David Klatzmann, Françoise Barré-Sinoussi, Jean-Claude Chermann y Luc Montagnier en *The New England Journal of Medicine*, vol. 311, n.° 20, págs. 1269-1273; 15 de noviembre de 1984.

IMMUNOLOGIC ABNORMALITIES IN THE ACQUIRED IMMUNODEFICIENCY SYNDROME (AIDS). A. S. Fauci en *Clinical Research*, vol. 32, n.° 5, págs. 491-499; diciembre, 1984.

EL SISTEMA ALCOHOLDESHIDROGENASA

THE COMPLETE AMINO ACID SEQUENCE OF THREE ALCOHOL DEHYDROGENASE ALLELOENZYMES FROM THE FRUITFLY *DROSOPHILA MELANOGASTER*. D. Thatcher en *Biochemical Journal*, vol. 187, págs. 875-886; 1980.

THE ALCOHOL DEHYDROGENASE POLYMORPHISM IN *DROSOPHILA* MELANOGASTER. W. van Delden en *Evolutionary Biology*, dirigido por M. K. Hecht, B. Wallace y G. T. Prance, vol. 15, págs. 187-222, Plenum Press; 1982.

ON THE EVOLUTIONARY RELATIONSHIPS OF *DROSOPHILA MELANOGASTER*. M. Ashburner, M. Bodmer y F. Lemeunier en *Developmental Genetics*, volumen 4, págs. 295-312; 1984.

CLONED *DROSOPHILA MELANOGASTER* ALCOHOL DEHYDROGENASE GENES ARE CORRECTLY EXPRESSED AFTER TRANSFECTION INTO *DROSOPHILA* CELLS IN CULTURE. C. Benyajati y J. F. Dray en *Proceedings of the National Academy of Sciences USA*, vol. 81, págs. 1701-1705; 1984.

EL TEOREMA ENORME

ON THE STRUCTURE OF GROUPS OF FINITE ORDER. Richard Brauer en *Proceedings of the International Congress of Mathematicians*, vol. 1; 1954.

FINITE GROUPS. Daniel Gorenstein. Chelsea Publishing Co., 1980.

FINITE SIMPLE GROUPS: AN INTRODUCTION TO THEIR CLASSIFICATION. Daniel Gorenstein. Plenum Press, 1982.

THE FRIENDLY GIANT. Robert L. Griess, Jr., en *Inventiones Mathematicae*, vol. 69, n.° 1, págs. 1-102; 1982.

LA ALIMENTACION EN CHINA

AGRICULTURE IN CHINA'S MODERN ECONOMIC DEVELOPMENT. Nicholas R. Lardy. Cambridge University Press, 1983.

THE BAD EARTH: ENVIRONMENTAL: DEGRADATION IN CHINA. Vaclav Smil. M. E. Sharpe, Inc., 1984.

RURAL DEVELOPMENT IN CHINA. Dwight Perkins y Shahid Yusuf. A World Bank Publication, The Johns Hopkins University Press, 1984.

PLANOS DEL TEMPLO DE APOLO EN DIDYMA

VORARBEITEN ZU EINER TOPOGRAPHIE VON DIDYMA. Klaus Tuchelt. Deutsches Archäologisches Institut, Instanbuler Mitteilungen, Beiheft 9, Verlag Ernst Wasmuth, Tübingen, 1973.

GREEK ARCHITECTS AT WORK: PROBLEMS OF STRUCTURE AND DESIGN. J. J. Coulton. Cornell University Press, 1977.

JUEGOS DE ORDENADOR

APPLICATIONS OF NUMBERED UNDIRECTED GRAPHS. Gary S. Bloom y Solomon W. Golomb en *Proceedings of the IEEE*, vol. 65, n.° 4, págs. 562-570; abril, 1977.

GOLOMB'S GRACEFUL CURVE. Martin Gardner en *Wheels, Life and Other Mathematical Amusements*. W. H. Freeman and Company, 1983.

VARIATIONS IN THE ROTATION OF THE EARTH. W. E. Carter, D. S. Robertson, J. E. Pettey, B.D. Tapley, B.E. Schutz, R. J. Eanes y Miao Lufeng en *Science*, volumen 224, número 4652, páginas 957-961; 1 de junio de 1984.

TALLER Y LABORATORIO

MULTIPLE IMAGES IN PLANE MIRRORS. Thomas B. Geenslade, Jr., en *The Physics Teacher*, vol. 20, págs. 29-33; enero, 1982.

REFLECTIONS IN A POLISHED TUBE. Laurence A. Marschall y Emma Beth Marschall en *The Physics Teacher*, vol. 21, pág. 105; febrero, 1983.

THROUGH THE KALEIDOSCOPE. Cozy Baker. Beechcliff Books, 100 Severn Avenue, Suite 605, Annapolis, Md. 21403; 1985.